Über die Zusammenhänge zwischen Grundfrequenz und Vokalhöhe: Evidenzen aus longitudinalen Altersstimmenstudien, Perturbations- und Vokalerkennungsexperimenten

Ulrich Reubold

München 2012

Über die Zusammenhänge zwischen Grundfrequenz und Vokalhöhe: Evidenzen aus longitudinalen Altersstimmenstudien, Perturbations- und Vokalerkennungsexperimenten

Ulrich Reubold

Inauguraldissertation
zur Erlangung des Doktorgrades der Philosophie
am Department II der Fakultät 13
(Sprach- und Literaturwissenschaften)
der Ludwig-Maximilians-Universität München

vorgelegt von
Ulrich Reubold
aus Mannheim

München, den 11.10.2011

Erstgutachter: Professor Dr. Jonathan Harrington
Zweitgutachter: PD Dr. habil. Phil Hoole
Tag der mündlichen Prüfung: 31. Januar 2012

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

Zusammenfassung

Diese Dissertation geht von einem Zusammenhang zwischen der Grundfrequenz und der Perzeption von Vokalen, speziell der Höhe von Vokalen, aus – wie viele Vorgängerstudien auch – und diskutiert Konsequenzen, die sich aus diesem Umstand ergeben; außerdem führt sie neue Evidenzen an, dass unter bestimmten Bedingungen die Grundfrequenz auch zur Produktion von Vokalhöhendistinktionen aktiv variiert werden kann.

In einer longitudinalen Studie wurden Aufnahmen aus mehreren Jahrzehnten, die von den selben britischen Sprechern stammten und auf Gleichwertigkeit der Kommunikationssituation kontrolliert worden waren, daraufhin untersucht, wie sich Alterungsprozesse in erwachsenen Sprechern auf die mittlere Grundfrequenz und die Formanten *F1*, *F2* und *F3* im Neutrallaut Schwa, bzw. auf die als äquivalent hierzu festgestellten gemittelten Formantwerte in allen stimmhaften Signalanteilen auswirken. Die Grundfrequenzen von Frauen werden als mit dem Alter fallend beschrieben, während Männer eine zunächst absinkende, später ansteigende Grundfrequenz aufweisen. Der zweite Formant ändert sich nur marginal, und auch *F3* weist keine über alle Sprecher konsistenten, signifikanten Änderungen auf. Im Gegensatz hierzu ändert sich *F1* mit zunehmendem Alter deutlich, und zwar bei den meisten Sprechern in die selbe Richtung wie die Grundfrequenz. In Daten eines Sprechers und einer Sprecherin, die in kurzen Abständen regelmäßig über ein halbes Jahrhundert hinweg aufgenommen worden waren, wird eine deutliche Kovariation des ersten Formanten mit der Grundfrequenz deutlich, wobei der Abstand zwischen *F1* und Grundfrequenz auf einer logarithmischen Skala auch über Jahrzehnte hinweg relativ invariant bleibt.

Die Hypothese hierzu ist, dass altersbedingte Formantänderungen weniger auf physiologisch bedingte Änderungen in den Abmessungen des Ansatzrohrs zurückzuführen seien, sondern auf eine kompensatorische Anpassung des ersten Formanten als Reaktion auf eine Perturbation des Vokalhöhenperzepts, welche hervorgerufen wird durch die (physiologisch bedingten) Grundfrequenzänderungen. Diese Hypothese schließt mit ein, dass das Vokalhöhenperzept der Sprecher/Hörer durch den *in Relation zu f0* zu beurteilenden ersten Formanten bestimmt ist.

Um diese letzte Schlussfolgerung weiter zu testen, wurden deutsche Sprecher in zwei Experimenten in Quasi-Echtzeit einem akustisch verändertem auditorischen *feedback* ausgesetzt, und ihre akustischen Daten untersucht. Beide Perturbationen hatten das Ziel, das Vokalhöhenperzept (direkt oder indirekt) zu beeinflussen:

Für eine Perturbation des ersten Formanten kompensierten die Sprecher mit einer *F1*-Produktion in Gegenrichtung zur Perturbation. Gleichzeitige Änderungen der produzierten

Grundfrequenz sind teilweise als automatisch eintretende Kopplungseffekte zu deuten; unter bestimmten Bedingungen scheinen manche Sprecher jedoch *f0* unabhängig von *F1* aktiv zu variieren, um die intendierte Vokalhöhe zu erreichen.

Bei einer Perturbation der Grundfrequenz variieren einige Sprecher den ersten Formanten dergestalt, dass zu vermuten ist, dass der aufgrund nur partiell durchgeführter *f0*-Kompensation weiterhin gegenüber den unperturbierten Werten veränderte *F1-f0*-Abstand das Vokalhöhenperzept beeinflusste, was zu einer kompensatorischen Gegenbewegung in Form einer Vokalhöhenvariierung führte.

Ein Perzeptionsexperiment mit ausschließlich durch Grundfrequenzvariierung beeinflussten Kontinua zwischen vorderen halb-geschlossenen und geschlossenen Vokalen in Wörtern gleichen Kontexts, welche in Trägersätze eingebettet präsentiert wurden, ergab, dass die Grundfrequenzvariation nur etwa bei der Hälfte der deutschen Hörer das Vokalperzept beeinflusste. Das *vokalintrinsische* Merkmal wird aber trotz des störenden Einflusses *extrinsischer* Faktoren genutzt, und auch trotz der intonatorischen Funktion der Grundfrequenz.

Die durch Ergebnisse von Untersuchungen zur Intrinsischen Grundfrequenz im Deutschen motivierte Hypothese, dass deutsche Hörer den *F1-f0*-Abstand als Vokalhöhenmerkmal in stärkerem Ausmaß in einem Kontinuum zwischen ungespannten Vokalen nutzen, als in einem Kontinuum zwischen gespannten Vokalen, konnte nicht bestätigt werden.

Generell liefern alle drei experimentellen Teile dieser Dissertation weitere Evidenz dafür, dass – zumindest in den vergleichsweise vokalhöhenreichen Sprachen Englisch und Deutsch – viele, aber eben nicht alle Sprecher/Hörer zur Vokalhöhenperzeption und -produktion neben *F1* auch die Grundfrequenz nutzen.

Kapitel 1

Einführung

1.1 Variabilität gesprochener Sprache

Schon kurz nachdem Phonetiker begannen, sich Gedanken über die Relation des akustischen Signals und der Perzeption der Sprache zu machen, wurde die Rolle von Prominenzen im Spektrum des Sprachsignals, die, über die Zeit, also im Sonagramm betrachtet, als Bänder auftreten und *Formanten* genannt werden, klar. Es wurde sehr schnell deutlich, dass es genügt, *steady-state*-Versionen, also über die Zeit invariante Ausprägungen der zwei niedrigsten Formanten zu präsentieren, um ein Vokalperzept hervorzurufen, und dass eine Variation der Lage dieser beiden ersten Formanten (*F1* bzw. *F2* genannt), die Wahrnehmung verschiedener Vokalkategorien hervorruft (Delattre, Liberman, Cooper & Gerstman, 1952) (obschon nicht unerwähnt bleiben soll, dass erste resonanzbasierte Vokalsynthesen bis zu von Helmholtz (1865) zurückzuverfolgen sind). Man kann durch eine solche Variierung der Lage der Mittenfrequenzen von *F1* und *F2* den gesamten Vokalraum abdecken, also alle Vokale (solange es sich nicht um nasalierte oder rhotizierte handelt) synthetisieren; *F1* ist hierbei invers mit der Zungenhöhe korreliert, und *F2* mit der Zungenlage. Allerdings – das anregende Quellsignal, also *f0*, wurde konstant gehalten, andere Parameter gab es gar nicht – handelt es sich bei Experimenten solcher Art um Einschätzungen der Vokale *einer* „Stimme“ . Betrachtet man verschiedene Stimmen, also die Daten verschiedener Sprecher – so war ebenso bald klar geworden – erhält man zwar, wie in dem „Klassiker“ phonetischer Literatur, Peterson und Barney (1952), Vokaltokens, die von menschlichen Hörern zu einem hohen Prozentsatz korrekt identifiziert werden können („korrekt“ im Vergleich mit den intendierten Äußerungen), deren akustische Korrelate aber eine hohe Variabilität aufweisen, die natürlich eine Zwischen-Sprecher-Variabilität und hierbei vor allem eine Variabilität zwischen den Geschlechtern und zwischen Erwachsenen und Kindern ist, so dass die akustischen Daten, abgebildet auf einer Ebene mit den Dimensionen *F1* (für die Vokalhöhe) und *F2* (für die Lage der Zunge), für die einzelnen Vokalkategorien eine starke Streuung und v.a. Überlappungen mit anderen Kategorien aufweisen. Zwei identische *F2-F1*-Paare könnten also, wenn sie von unterschiedlichen Sprechern stammen, zwei unterschiedliche Vokalkategorien repräsentieren, und zwei tokens der gleichen Vokalkategorie,

insbesondere (aber nicht nur) wenn sie von unterschiedlichen Sprechern stammen, können stark divergierende *F1*- und *F2*-Werte aufweisen. Wie oben beschrieben, fällt es Hörern aber nicht allzu schwer, diese Vokale korrekt zu kategorisieren – offenbar wenden Hörer also eine Normalisierung an.

1.2 Quellen der Variation

Wie schon Ladefoged und Broadbent (1957) feststellten, überträgt ein Sprachsignal mehrere Information auf einmal, oder, anders ausgedrückt, die Variation in einem Sprachsignal speist sich aus mehreren Quellen. Natürlich ist im Sprachsignal *linguistische* Information kodiert, d.h. dass Teile der Variation den akustischen Korrelaten phonologischer Merkmale geschuldet sind; die restliche Variation ist sprecherbedingt, wobei sich diese Variationsquelle wiederum aufteilen lässt in *sozio-linguistische* Variation, die den Sprecher als Teil einer bestimmten Gruppe von Sprechern ausweist, z.B. als Sprecher einer bestimmten Varietät einer bestimmten Sprache, und in *persönliche* Variation, die auf die indiosynkratischen Eigenschaften des Sprechers zurückzuführen ist. Unter diese idiosynkratischen Merkmalen sind in erster Linie die physiologischen Unterschiede zwischen den Sprechern zu zählen, doch diese reichen, wie Traunmüller (2005) meint, nicht ganz aus, um die *persönliche* Variation zu erklären, da noch die *expressive* Komponente hinzutritt, also die Veränderungen der Stimme in verschiedenen situativen Kontexten, oder zu verschiedenen emotionalen Zuständen des Sprechers.

Nun ist die Normalisierung, die menschliche Hörer vornehmen, ein bis heute nicht zur Gänze aufgeklärter Prozess; man weiß aber, dass die Normalisierung nicht nur auf den im akustischen Signal kodierten Informationen beruhen muss, sondern auch ein aktiver Prozess ist (Mullenix, Pisoni & Martin, 1989), der auch von visuellen Eindrücken und bloßen *Erwartungen* beeinflusst werden kann. So zeigten z.B. Johnson, Strand und D'Imperio (1999), dass die visuelle Präsentation von weiblichen oder männlichen Gesichtern die Aufteilung ein und desselben Kontinuums beeinflussen kann, und selbst die bloße Vorstellung, man höre eine weibliche oder männliche Stimme beeinflusst die Antworten auf ein Kontinuum. Dies betrifft nicht allein die *Erwartungen* der Hörer bezüglich „weiblicher“ und „männlicher“ Stimmen, also der Bewertung des *organischen* Teils persönlicher Information, sondern gilt auch für die Bewertung *sozio-linguistischer* Variation, denn die perzeptuelle Aufteilung von Kontinua lässt sich durch die wahrgenommene Zugehörigkeit der Sprecher zu einer Dialektgruppe (Niedzielski, 1999) oder durch die wahrgenommene Zugehörigkeit zu einer Altersgruppe (Drager, 2011), aber sogar durch mit dem Sprachsignal scheinbar nichts zu tun habenden Äußerlichkeiten beeinflussen, wie der Anwesenheit von Spielzeugkängurus, -koalas, bzw. -kiwis während eines Experiments, die offenbar beeinflussten, ob der Sprecher eher für einen Australier (Kängurus und Koalas) oder einen Neuseeländer (Kiwis) gehalten wird (Hay & Drager, 2010), was Versuchspersonen dazu veranlasst, Stimuli verschieden zu kategorisieren.

Diese Erkenntnisse zeigen aber auch, dass Hörer rein aus dem akustischen Sprachsignal die oben genannten Informationsarten voneinander unterscheiden können, da sie nicht nur

zwischen Sprecherstimme und linguistischem Inhalt unterscheiden können, sondern eben offenbar auch Konzeptualisierungen über bestimmte Relationen zwischen dem Sprecher und dessen akustischem Output entwickelt haben, und so erkennen können, dass der gegenwärtig gehörte Sprecher einer bestimmten sozio-linguistisch zu definierenden Gruppe angehört, welches Geschlecht er/sie hat, wie alt er/sie ist, in gewissen Rahmen auch dass er/sie ein bestimmter Sprecher ist (Sprechererkennung, van Dommelen (1990)), und in welchem emotionalem Zustand ein Sprecher ist (Imaizumi et al., 1997); auch Kommunikationsbedingungen wie die Entfernung zwischen Sprecher und Adressat des Gesprochenen sind für den Hörer aus dem akustischen Signal abschätzbar (Eriksson & Traunmüller, 2002).

1.3 Alters- und geschlechtsbedingte Kovariation mehrerer Parameter

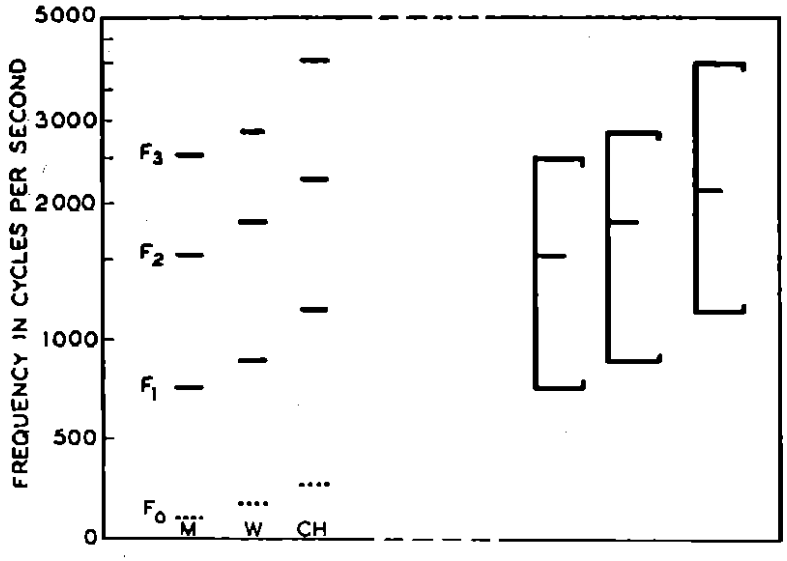

FIG. 5. Formant frequencies for [æ] (had) as spoken by a man, a woman and a child.

Abbildung 1.1: *f0 und F1-F3 bei* [æ] *für einen Mann, eine Frau, und ein Kind, aus Potter und Steinberg (1950, Seite 812), mit originaler Abbildungsbeschriftung.*

Abbildung 1.1 zeigt am Beispiel eines Vokals, gesprochen von einem Mann, einer Frau, und einem Kind, die unterschiedlichen Formantlagen. Diese Geschlechts- und Altersunterschiede kommen, wie interindividuelle Unterschiede überhaupt, hauptsächlich durch die unterschiedliche Größe und Form des Ansatzrohrs zustande. So zeigen anatomische Daten (Goldstein, 1980; Fitch & Giedd, 1999) nicht ein gleichmäßiges Wachstum des gesamten Sprechapparats, sondern eine Änderung der Proportionen zwischen oralem und pharyn-

galem Teil des Ansatzrohrs, wobei der pharyngale Raum deutlicher wächst, auch und vor allem durch die zweite Absenkung des Kehlkopfs während der Pubertät, die hauptsächlich bei männlichen Heranwachsenden zu finden ist[1]. Schon Fant (1975) hatte aufgrund von Nicht-Uniformitäten auf akustischer Ebene, also der Tatsache, dass man nicht einfach Formantdaten weiblicher Sprecher mit einem Faktor multiplizieren kann, um zu Werten von männlichen Sprechern zu gelangen, vermutet, dass bei Männern proportional der pharyngale Raum größer sei als bei Frauen, was durch diese anatomischen Daten größtenteils bestätigt wird (siehe aber auch Turner, Walters, Monaghan und Patterson (2009), dessen Autoren aus den Daten in Fitch und Giedd (1999) keinen Geschlechtsdimorphismus bezüglich der Proportionen zwischen oraler und pharyngaler Länge erkennen können).

Doch, wie auch schon die Abbildung 1.1 andeutet, gibt es offenbar gleichzeitige Änderungen des Kehlkopfes und damit der Grundfrequenz. Der Kehlkopf wächst, und Länge und Masse der Stimmlippen nehmen – neben der generellen Vergrößerung – zu (Hirano, Kurita & Nakashima, 1983), und zwar wiederum mehr bei Männern als bei Frauen. Wir haben es also mit einer gewissen Kovariation von Larynxgröße und Vokaltraktgröße, bzw. in akustischen Maßen, von durchschnittlicher Grundfrequenz und Formanten zu tun.

Diese Kovariation über die Sprechergruppen Männer, Frauen, und Kinder, zeigt auch Abbildung 1.2. Man sieht anhand der Daten aus Peterson und Barney (1952), dass Männer die tiefsten Grundfrequenzen, Frauen die zweittiefsten, und Kinder die höchten haben, sowie, dass sowohl *F1*, *F2* und *F3* umso höher sind, je höher *f0* ist. Dies gilt auch für das Korrelat der Vokaltraktlänge, nämlich das geometrische Mittel der ersten drei Formanten. Das gleiche Maß wurde auch in Assmann et al. (2008) verwendet, und deren Daten sind ebenfalls vergleichbar. Dort wurde nur der Altersausschnitt von 5 bis 18 Jahren untersucht, und dort war die Korrelation zwischen dem geometrischen Mittel der Formanten und der Grundfrequenz für männliche Versuchspersonen ausgeprägter als bei weiblichen. Weitere akustische Daten zur Entwicklung im Übergang vom Kindes- zum Erwachsenenalter sind in Lee, Potamianos und Narayanan (1999) und der Reanalyse der gleichen Daten von 436 Heranwachsenden im Alter von 5 bis 17 und weiterer 57 Erwachsener in Whiteside (2001) zu finden.

Nun mögen diese Befunde über eine Tendenz zur Kovariation von Grundfrequenz und allgemeiner Formantlage, bzw. von Kehlkopf- und Ansatzrohrgröße ausgesprochen trivial erscheinen; man muss sich aber klarmachen, dass Hörer ständig Beispiele von Sprache wahrnehmen, die von Stimmen stammen, bei denen *in der Regel* die Grundfrequenz *und* die Formanten tief (bzw. hoch) sind. Wird von dieser „Daumenregel“ abgewichen, sinken auch prompt die Erkennungsraten, wie beispielsweise Gottfried und Chew (1986) für einsilbige Wörter, produziert von einem Countertenor, also einem Sänger, der das Falsettregister einsetzt, um im „weiblichen“ Grundfrequenzbereich zu singen, dabei aber natürlich sein „männliches“ Ansatzrohr behält, zeigten; eine Analyse der Vokalisierungen des Countertenors zeigte hierbei übrigens einen etwa 10%igen Anstieg des ersten Formanten bei Ver-

[1] die erste Absenkung findet in sehr frühem Alter statt und ermöglicht erst die Artikulation; die hohe Lage bei Säuglingen ermöglicht hingegen gleichzeitiges Atmen und Nahrungsaufnahme; vergleiche zu diesem Thema D. E. Lieberman, McCarthy, Hiiemae und Palmer (2001).

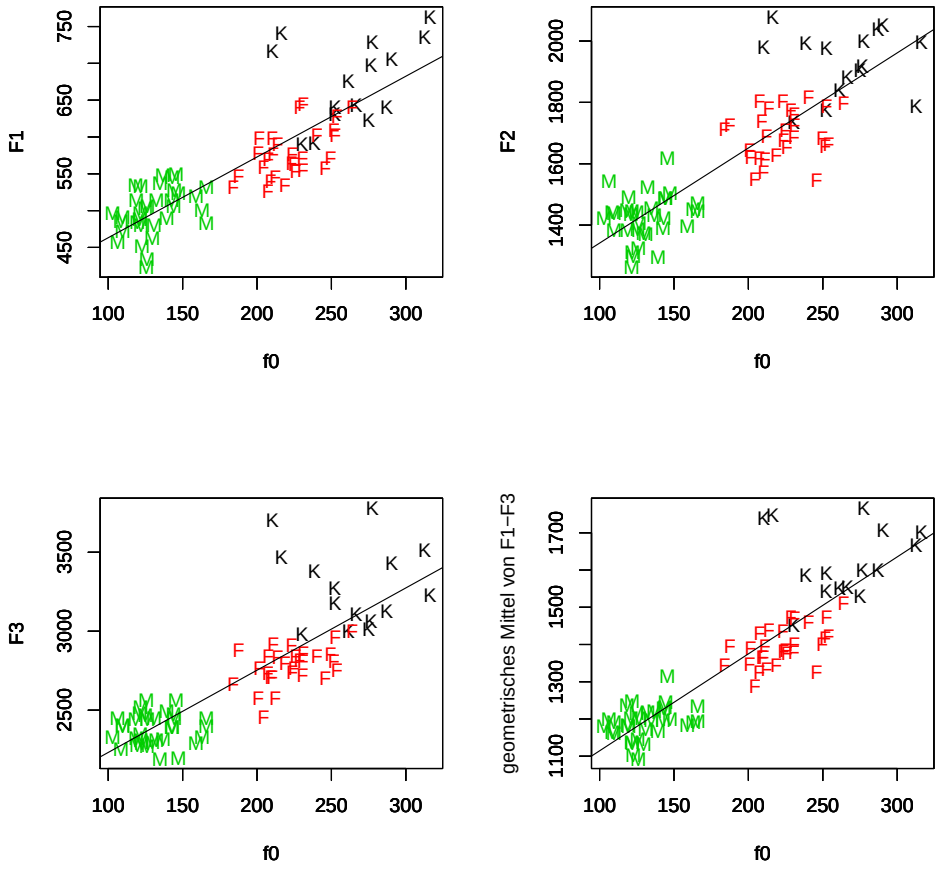

Abbildung 1.2: *Männer (M, grün), Frauen (F, rot), und Kinder (K, schwarz): F1-F3 und das geometrische Mittel von F1-F3 als Funktion von f0. Daten der 76 Sprecher aus Peterson und Barney (1952), gemittelt über alle 10 in dieser Studie untersuchten Vokale in* /hʊd/*-Kontext. Das geometrische Mittel gilt gemäß Assmann et al. (2008) als Korrelat der Vokaltraktlänge, und weist bei linearer Korrelation hier den größten R^2-Wert auf (0,78). Weitere Werte sind: für F1: 0,72; für F2: 0,74; für F3: 0,69. (jeweils p<0,001)*

dopplung der Grundfrequenz. Auch die Sprache von Tauchern, die ein Helium-Sauerstoff-Gemisch atmen, was zwar kaum die Stimmlippenschwingung und damit die Grundfrequenz, sehr wohl aber wegen der veränderten Schallgeschwindigkeit im weniger dichten Medium Helium die Resonanzen und damit die Formanten verändert, wird schwer verständlich und muss deshalb sogar in der Regel über technische Mittel, die die Formanten wieder verschieben, zurücktransformiert werden (Golden, 1966; Mendel, Hamill, Crepeau & Fallon, 1995). Schon Potter und Steinberg (1950) stellten fest, dass die Vokalqualität sich ändert, wenn man ein „männliches“ Formantmuster mit einer „kindlichen “ Grundfrequenz anregt. Lehiste und Meltzer (1973) präsentierten die Formantmuster männlicher, weiblicher

und kindlicher Vokale kombiniert mit den Grundfrequenzen dieser drei Sprechergruppen und zeigten – zumindest in manchen der „unpassenden“ Kombinationen – ein Absinken der Erkennungsraten. Nur bei den männlichen Formanten wurden ähnliche Erkennungsraten erzielt, unabhängig davon, ob diese mit männlichen oder weiblichen Grundfrequenzen präsentiert wurden – in beiden Fällen war die Rate allerdings mit unter 77% relativ gering. Ryalls und Lieberman (1982) zeigten ebenfalls eine Abnahme der Erkennungsraten bei einem Mismatch von Formanten und Grundfrequenzen; die Verschlechterung der Erkennungsraten war aber geringer ausgeprägt, wenn eher hohe Formanten mit eher tiefen Grundfrequenzen präsentiert wurden, was darauf schließen lässt, dass die Dichte der Harmonischen der Grundfrequenz für diese Effekte eine Rolle spielt (zum *sampling* des Spektrums durch die Harmonischen der Grundfrequenz siehe später etwas mehr). Vergleichbares finden auch Assmann und Nearey (2008).

Grundsätzlich ist die Änderung der Formantfrequenzen vom Kindes- zum Erwachsenenalter natürlich geringer als jene der Grundfrequenz, wie auch Abbildung 1.2 zeigt. Einer Halbierung der Grundfrequenz entspricht keineswegs eine Halbierung der Formantfrequenzen, sondern nur ein Bruchteil davon. So sinken denn Erkennungsraten auch relativ schnell, wenn man, wie Chiba und Kajiyama (1941) (zitiert nach Assmann und Nearey (2008)), ein Sprachsignal so verändert, dass sowohl für die Grundfrequenz als auch die Formanten der selbe Faktor angewandt wird, wie es der Fall ist, wenn man eine Aufnahme mit veränderter Geschwindigkeit abspielt. Die Verständlichkeit bleibt hierbei für Faktoren zwischen 0.8 bis 1.5 aber durchaus einigermaßen erhalten.

1.4 *f0* und höhere Formanten als Anhaltspunkt?

Es erscheint dennoch wegen der offensichtlichen Kovariation der Grundfrequenz mit den Formantlagen also naheliegend, dass Hörer die Grundfrequenz als (zumindest groben) Anhaltspunkt für eine Normalisierung der variablen Formanträume nehmen könnten. Noch näher, so mag es zunächst scheinen, liegt es jedoch, die höheren Formanten als Anhaltspunkt für eine Abschätzung der Formantlagen heranzuziehen, da ja alle Formanten *einem* Ansatzrohr entstammen. So spielt *F3* für die Distinktion verschiedener Vokalphoneme nur selten eine Rolle, am ausgeprägtesten sicherlich für rhotizierte Vokale (vgl. auch die Diskussion in Fujisaki und Kawashima (1968)). Davon abgesehen, bleibt *F3* in den Daten aus Peterson und Barney (1952) zumindest bei Männer- und Kinderstimmen relativ fixiert (vgl. die Tabelle in Kent (1993, Seite 103)), sinkt aber bei gerundeten Vokalen, was aber eher ein bestenfalls sekundärer akustischer Cue ist, der durch die durch die Rundung entstandene Vokaltraktlängung und der dadurch bedingten Absenkung aller Formanten zustandekommt; die Behauptung, dass dieser Effekt, obschon in den gemessenen Daten durchaus deutlich, eher als sekundär wahrgenommen wird, ist darin begründet, dass man durchaus beispielsweise ein /iː-uː/-Kontinuum erzeugen kann, indem man nur *F2* variiert, *F3* aber unverändert lässt (z.B. in Harrington, Kleber und Reubold (2008)). So sollte, trotz einer gewissen Variation über die Vokalphoneme hinweg, *F3* eine Abschätzung der Dimensionen des Vokalraums ermöglichen, zumal, wie schon Peterson (1952) feststellte, *F3* mit den zwei

ersten Formanten recht gut korreliert. Fujisaki und Kawashima (1968) zeigten aber eher gering ausgeprägte Effekte für eine *F3*-Variation auf die Aufteilung stimmhaft angeregter Vokalkontinua, dafür allerdings stärkere, wenn die Vokale durch Rauschen angeregt wurden. Auch Slawson (1968) zeigte einen im Vergleich zu Effekten, die durch Variation der Grundfrequenz ausgelöst wurden, geringen Effekt einer Variation von *F3*. Nur für vordere, nicht aber für hintere Vokale fand Johnson (1989b) einen Effekt von *F3*, so dass er diesen Effekt der spektralen Integration von *F2* und *F3* bei Abständen von unter drei Bark, also dem sogenannten *effektiven F2* oder *F2'* (siehe hierzu z.B. Chistovich und Lublinskaya (1979)) anrechnete. Generell scheint also der Einfluss von *F3* auf die Wahrnehmung von Stimuli relativ beschränkt zu sein, wenn denn eine stimmhafte Anregung involviert ist, denn die Grundfrequenz alleine scheint, wie wir gleich sehen werden, stärkere Einflüsse auszuüben. Allerdings spiegelt sich die oben erwähnte triviale Feststellung, dass zumeist Kehlkopf- und Ansatzrohrgröße kovariieren, darin wider, dass die stärksten Effekte gefunden werden, wenn sowohl *f0* als auch *F3* variiert werden, wie Fujisaki und Kawashima (1968) zeigten.

Wie angedeutet, sind Effekte, die nur auf der Variation der Grundfrequenz beruhen, als stärker zu betrachten als jene von *F3* oder überhaupt *höherer* Formanten (also jener ab *F3*); *f0* ist also offenbar ein stärkerer Cue (Nearey, 1989), der zur Normalisierung genutzt werden kann. Schon R. L. Miller (1953) präsentierte die gleichen Formanten mit zwei Grundfrequenzlevels, bei 120 Hz und bei 240 Hz, und fand Verschiebungen der Grenzen der Vokalkategorien. Fant, Carlson und Granström (1974) fanden, dass die selben *F1-F2*-Werte verschiedene Vokalperzepte hervorrufen, abhängig davon, mit welcher Grundfrequenz sie präsentiert werden. Auch Fujisaki und Kawashima (1968) fanden Verschiebungen der Kategorisierung von Vokalen, die besonders auf der *F1*-Dimension besonders ausgeprägt waren, oder in anderen Worten, auf der Vokalhöhendimension. Eine Perzeptionsstudie in Reinholt Petersen (1986), in der ein Kontinuum von Dänisch /b̥uːðə/ bis /b̥oːðə/ in drei Versionen, nämlich mit drei parallelen Grundfrequenzverläufen, die jeweils 6 Hz auseinander lagen (und damit um den von Reinholt Petersen (1978) festgestellten Unterschied an *intrinsischer Grundfrequenz* zwischen Dänisch [uː] und [oː]), präsentiert wurde, zeigte, dass die Unterschiede in den Grundfrequenzen in der Lage waren, die Antwortkurven signifikant zu beeinflussen. Auch Hirahara und Kato (1992) fanden, dass verschiedene *f0*-Levels Vokalperzepte beeinflussen, wobei es der Fall ist, dass eine durch eine *f0*-Änderung verursachte Kategorienänderung immer eine Vokal*höhen*-Veränderung ist. Sie zeigen aber auch dennoch einen gewissen Einfluss auf die Kategorisierung entlang der traditionell mit der Zungenlage assoziierten *F2*-Achse. Dass die Grundfrequenz, wenn sie einen Einfluss auf Vokalkategorisierung ausübt, hauptsächlich einen Effekt auf die Vokalhöhe hat, wird auch in den Arbeiten Traunmüllers (beginnend mit Traunmüller (1981)) und zahlreicher nachfolgender Arbeiten auch anderer Autoren gezeigt, die auf einen Zusammenhang zwischen der Grundfrequenz und dem ersten Formanten hinweisen. Wir wollen diese Arbeiten etwas später besprechen und zunächst auf eine weitere „natürliche“ Variationsquelle aufmerksam machen, die – neben eventuell auftretenden perzeptuellen Gründen – einen relativen Einfluss der Grundfrequenz auf die Vokalkategorisierung, und hierbei speziell die Vokal*höhen*kategorisierung, sehr wahrscheinlich macht: die sogenannte *Intrinsische Grund-*

frequenz, genauer, die vokalintrinsische Grundfrequenz.

1.5 Kovariationen innerhalb eines Sprechers – Intrinsische Grundfrequenz

Bislang wurde hier so getan, als wären die unterschiedlichen Vokale *eines* Sprechers hauptsächlich durch die von ihm produzierten Unterschiede in den ersten beiden Formanten charakterisiert. Es gibt aber gleich mehrere Phänomene, die mit diesen Änderungen kovariieren, wobei diese Kovariationen in der Regel Tendenzen sind, die nur dann stabil aufzudecken sind, wenn andere Parameter unverändert bleiben, also z.B. wenn unterschiedliche Vokale im gleichen Kontext auftauchen.

Zwar ist es oft der Fall, dass mehrere Parameter mit bestimmten artikulatorischen Gesten kovariieren und manche als primäre, manche als sekundäre Cues zu einem bestimmten Perzept beitragen können, so wie z.B. Vokaldauer zum Perzept einer Vokalqualität beitragen kann (Lehiste & Meltzer, 1973), wobei diese Daueunterschiede darauf zurückgeführt werden können, dass es länger dauert, einen *offenen* Vokal zu produzieren, als einen *geschlossenen*, da die Artikulatoren für einen *offenen* Vokal einer weitere Strecke zurückzulegen haben. Andererseits gilt dies als stabiler Effekt eben nur im gleichen Kontext, da die gleichen Parameter oft zusätzlich mit *suprasegmentaler* Variation kovariiert. Als ein solches Beispiel kann gelten, dass z.B. Intensitätsunterschiede neben relativen *f0*- und Daueunterschieden kommunikative Funktionen wie Prominenzmarkierung erfüllen (Kochanski, Grabe, Coleman & Rosner, 2005), obschon eine relativ starke Variabilität an Intensitätsunterschieden schon „vorgegeben“ ist, z.B. durch Zugehörigkeit zu einer Phonemklasse, denn Vokalhöhe und Intensität sind negativ miteinander korreliert (und damit *F1* positiv mit Intensität). Turk und Sawusch (1996) zeigen aber auf, dass Intensitätsunterschiede *unabhängig von anderen Parametern* (wie z.B. Dauer) nur wenig linguistisch Relevantes auf segmenteller Ebene beizutragen vermögen.

All diese Kovariationen sind eigentlich nur Nebenprodukte der Produktion – so sind *offenere* (bzw. *tiefere*) Vokale im Mittel länger als *geschlossene* (oder anders: *hohe*), da es – wie bereits erwähnt – eben länger dauert, den Kiefer weiter zu öffnen, und außerdem sie sind aus akustischen Gründen wegen der größeren Öffnung mit höherer Intensität verknüpft. Es sind also biomechanische oder akustische Kopplungen, die diese Effekte bedingen, und all diese Effekte folgen einem gewissen Automatismus. Dennoch haben diese gekoppelten „Nebenprodukte“ , wie wir sie oben bezeichnet haben, Konsequenzen auf das akustische Sprachsignal. Stevens und Keyser (2010) beschreiben[2], wie diese Konsequenzen der „Nebenprodukte“ der Artikulation die *Salienz*, also die Auffälligkeit eines Reizes innerhalb seines Kontextes, erhöhen können, und unterscheiden folgerichtig beide artikulatorischen Gestenarten in *feature defining* und *feature enhancing gestures*; desweiteren stellen sie fest, dass durch starke koartikulative Effekte die merkmalsdefinierenden Gesten oft stärker betroffen sind als die merkmalsverstärkenden, wobei die akustischen Konsequen-

[2]Sie tun dies, die Autorennamen lassen es vermuten, im größeren Rahmen der Quantaltheorie

zen der letzteren dem Hörer ermöglichen, den intendierten Laut zu erkennen; die oben so genannten „Nebenprodukte“ spielen also offenbar eine wichtige Rolle für Produktion und Perzeption der Sprache.

Eines dieser „Nebenprodukte“ ist die sogenannte *Intrinsische Grundfrequenz*, also der Umstand, dass *hohe* Vokale mit einer höhereren Grundfrequenz produziert werden als *tiefe* Vokale[3]. Dies bedeutet gleichzeitig, dass, da *F1* invers mit Vokalhöhe korreliert ist, *f0* und *F1* ebenfalls negativ korreliert sind.

Vielleicht mit der Ausnahme von Diehl und Kluender (1989) und Kingston (1992) bestreitet praktisch niemand, dass es sich bei der Intrinsischen Grundfrequenz um einen ebenfalls automatisch bedingten Kopplungseffekt handelt, wobei es nicht ganz geklärt ist, wie man sich diese Kopplung vorzustellen hat. Neben einer eher abzulehnenden (Ohala, 1973; Ewan & Ohala, 1979) akustischen Kopplung, wie sie nach dem Befund von Flanagan und Landgraf (1968), dass der erste Formant einen Einfluss auf die Schwingung der Stimmlippen auslöst, wenn *F1* und *f0* nur nahe genug sind, vorgeschlagen worden war, ist eine biomechanische Kopplung oft zur Erklärung herangezogen worden. Wie die Überblicke zu dieser Frage z.B. in Ohala (1973), Dyhr (1990), K. Honda (2004) und Hoole (2006) zeigen, gab es hierzu nicht eine, sondern mehrere aufeinanderfolgende Theorien, die teilweise wieder aufgegeben werden mussten; gemein ist diesen, dass angenommen werden muss, dass die Zunge über das Zungenbein mit dem Kehlkopf mechanisch gekoppelt ist und somit die Rotation zwischen dem Cricoid und dem Thyroid, und wiederum damit die Stimmlippenlänge und -spannung (die wiederum für die Frequenz der Stimmlippenschwingung entscheidend sind) beeinflusst werden kann, wobei dies vom Ausmaß der Kontraktion des genioglossus-Muskels abhängig ist. Es kommt uns hier *nicht* darauf an, die Art der biomechanischen Kopplung zu beschreiben, sondern lediglich, dass eine solche vorhanden ist, ja eigentlich vorhanden sein muss. Für diese These spricht nämlich auch, dass Intrinsische Grundfrequenz ein praktisch universales Phänomen ist (Whalen & Levitt, 1995), und auch in Sprachen vorkommt, wo sie eher hinderlich ist, also etwa in Tonsprachen (siehe zu dieser Frage auch Connell (2002), der möglicherweise ein Gegenbeispiel aufzuweisen hat, die Existenz von intrinsischer Grundfrequenz in Tonsprachen aber prinzipiell bestätigt).

Wir haben also gesehen, dass die Intrinsische Grundfrequenz ein (wahrscheinlich) universales Phänomen ist, das (wahrscheinlich) einer Kopplung (wahrscheinlich biomechanischer Art) entspringt, und dass das Ergebnis eine Grundfrequenz ist, die, gegeben, dass alles andere unverändert bleibt („all other things being equal“ (Ohala, 1973)), sich mit zunehmender Vokalhöhe in Gegenrichtung zum ersten Formanten bewegt; d.h. dass der Abstand zwischen beiden Maßen von beiden „Seiten“ aufeinander zukommend verringert wird. Da in dieser Betrachtungsweise die Unterschiede zwischen hohen und tiefen Vokalen deutlicher

[3]Higgins, Netsell und Schulte (1998) zeigen nicht nur vokalinherente Effekte für die Grundfrequenz, also Intrinsische Grundfrequenz, sondern auch vokalinherente Variation für subglottalen Luftdruck (höher bei /iː/ als bei /ɑː/), *end air flow* (Luftstrom im ausklingenden Vokal gemessen; ist geringer bei /iː/ als bei /ɑː/), *electroglottograph cycle width* (ebenfalls geringer bei /iː/ als bei /ɑː/) und in der *voice onset time* (VOT) für die adjazenten Plosive (VOT ist größer z.B. in [piː] als in [pɑː]), und führen alle diese Effekte auf eine Mischung aus mechanischer Kopplung sowie dadurch bedingt veränderter neuronaler, aber auch gelernter neuronaler Steuerung zurück.

werden als durch *F1* alleine, da man zwei Parameter hat, die sich gegenläufig aufeinander zu oder voneinander weg bewegen, kann man leicht auf den Gedanken verfallen, dass dieses Phänomen eine Bedeutung auf perzeptiver Ebene haben könnte. Geht man hiervon aus, muss man auch davon ausgehen, dass *f0* auch – neben seiner Bedeutung auf intonatorischen, also suprasegmenteller Ebene – auf segmenteller Ebene ein Perzept auslöst, das die Vokalhöhenwahrnehmung betrifft. Geht man wiederum hiervon aus, so kann man sogar soweit gehen, dass die Intrinsische Grundfrequenz praktisch ausschließlich aktiv (Diehl & Kluender, 1989; Kingston, 1992) zum Zwecke des *feature enhancements* eingesetzt wird, so dass eine Vokalhöhendistinktion durch eine absichtsvoll von Sprechern herbeigeführte Koordination *unabhängiger* Artikulatoren realisiert wird, und ein Automatismus durch Kopplungen gar nicht gegeben sei. Die Gegenposition, die vertritt, dass die intrinsische Grundfrequenz ausschließlich einer rein automatischen Kopplung entspringt, wird hauptsächlich von Whalen (Whalen & Levitt, 1995; Whalen, Gick, Kumada & Honda, 1999) repräsentiert. Die mittelnde Position zwischen diesen Extremem, die davon ausgeht, dass die intrinsische Grundfrequenz zumindest teilweise aktiv eingesetzt wird, und in diesem Fall umso stärker, je reicher ein Vokalinventar einer Sprache in der Vokalhöhendimension ist, wird beispielsweise von Hoole (2006) und Hoole und Honda (2011) vertreten.

Es ist umstritten, ob es Befunde dafür gibt, dass die intrinsischen Grundfrequenzeffekte in Sprachen mit reicher Vokalhöhendifferenzierung stärker ausgeprägt sind als in Sprachen mit weniger Vokalhöhenunterschieden. Whalen und Levitt (1995) fassen Untersuchungen zu 31 Sprachen aus 11 Sprachfamilien zusammen und finden unterschiedliche Ausmaße der intrinsischen Grundfrequenz, die aber laut den Autoren nicht in Zusammenhang mit der Größe des jeweiligen Vokalinventars zu bringen sind, sondern eher mit Artefakten der jeweiligen Untersuchungsmethode. Das (erst durch die Interpretation durch die Autoren entstandene) Fehlen eines Zusammenhangs zwischen der Vokalinventargröße und des Ausmaßes der intrinsischen Grundfrequenz ist für die Autoren ein Beleg dafür, dass die intrinsische Grundfrequenz eine automatische Konsequenz der Vokalproduktion ist und nicht aktiv vom Sprecher beeinflusst wird. Verhoeven und Van Hoof (2007) und v.a. Van Hoof und Verhoeven (2011) zeigen allerdings im Vergleich des vokalarmen Arabischen (Marokkanisches Arabisch mit 3 Vokalphonemen mit 2 Vokalhöhen, /i a u/) und des vokalreichen Niederländischen (Belgisches Niederländisch („Flämisch“) mit 12 monophtongalen, nicht-reduzierten Vokalphonemen /i ɪ e ɛ a ɑ o ɔ u y ʏ ø/), dass bei Muttersprachlern des Niederländischen intrinsische Grundfrequenz (*If0*) stärker ausgeprägt ist($\Delta If0_{i:-a:[L1=Niederländisch]} = 2.28$ Halbtöne) als bei Muttersprachlern des Arabischen ($\Delta If0_{i:-a:[L1=Arabisch]} = 0.74$ Halbtöne), auch wenn diese Niederländisch als L2 ($\Delta If0_{i:-a:[L2=Niederländisch]} = 0.83$ Halbtöne) sprechen. Da im Vergleich beider Varianten des Niederländischen, gesprochen von Muttersprachlern und von marokkanischen L2-Lernern, keine signifikanten Unterschiede bezüglich des F2-F1-Raumes gefunden wurden, gehen die Autoren davon aus, dass Muttersprachler des Niederländischen die intrinsische Grundfrequenz aktiv zur Kontrastverstärkung einsetzen, und ihre breitere intrinsische Grundfrequenz-Verteilung im Niederländischen nicht von Unterschieden in der Artikulation mit dadurch bedingter automatischer Beeinflussung der Grundfrequenz abhängig ist.

Fischer-Jørgensen (1990) zeigte, dass [±*gespannt*]-Vokalpaare trotz der tieferen Zun-

genposition ungespannter Vokale mit ähnlicher intrinsischer Grundfrequenz produziert werden, ein Befund, der bezüglich der vertikalen Zungenposition von Mooshammer, Hoole, Alfonso und Fuchs (2001) und Hoole und Mooshammer (2002) zusätzlich mittels Elektro-Magnetischer Midsaggitaler Artikulographie (EMMA) bestätigt werden konnte. Hierbei verhält es sich so, dass beispielsweise der ungespannte Vokal /ɪ/ eine tiefere Zungenposition aufweist als der nach traditioneller Beschreibung „tiefere“, gespannte Vokal /eː/, aber dafür eine vergleichbare intrinsische Grundfrequenz wie sein gespanntes Gegenstück, der Vokal /iː/. Anschaulicher ausgedrückt heißt dies, dass bezüglich der vertikalen Zungenposition $/i:/ > /e:/ > /I/$, während die Reihenfolge bezüglich intrinsischer Grundfrequenz $/i:/ > /I/ > /e:/$ (wobei oftmals $/i:/ \approx /I/$, siehe Fischer-Jørgensen (1990), aber auch Pape und Mooshammer (2004, 2006a)) ist, die intrinsische Grundfrequenz also die „traditionelle“ , auf impressionistische Beschreibung zurückgehende Vokal*höhe* besser abbildet als die vertikale Zungenposition. Mooshammer et al. (2001) zeigten desweiteren, dass bei einer Perturbation der Artikulation durch einen Beißblock die intrinsische Grundfrequenz in verstärktem Maße eingesetzt wird, was man so deuten kann, dass intrinsische Grundfrequenz wohl auch aktiv zur Kompensation eingesetzt wird. Man kann diesen Befund aber auch rein mechanisch motiviert beschreiben: Ohala und Eukel (1987) erhalten ebenfalls bei einem Beißblockexperiment das Ergebnis, dass intrinsische Grundfrequenz sich unter Beißblockperturbation verstärkt, deuten dies jedoch als weitere Evidenz für die tongue-pull-hypothesis. Da in beiden Fällen keine Elektromyographie-Daten für Kehlkopfmuskeln vorliegen, muss der Grund für die Erweiterung der vokalintrinsischen *f0*-Variation im Unklaren bleiben.

Der musculus cricothyroideus gilt als derjenige Muskel, der bei der aktiven Kontrolle der Grundfrequenz die Hauptrolle zu spielen scheint, und dessen Beitrag zur intrinsische Grundfrequenz im Sinne einer positiven Korrelation seiner Aktivität mit intrinsische Grundfrequenz für einige Sprachen gezeigt werden konnte (so Dyhr (1990) für Dänisch, Auteserre, Di Cristo und Hirst (1986) für Französisch, K. Honda und Fujimura (1991) für Englisch und Vilkman, Aaltonen, Raimo, Arajärvi und Oksanen (1989) für das Finnische). Whalen et al. (1999) versuchen jedoch, zu zeigen, dass dies nicht automatisch eine aktive Steuerung der intrinsische Grundfrequenz zu bedeuten hat. Mittels einer etwas fern von gewöhnlicher Sprachproduktion stehenden Aufgabe, isolierte Vokale zu bestimmten Grundfrequenzen zu produzieren, stellen sie vokalspezifische Unterschiede bezüglich der Korrelation von *f0* und Aktivität der pars recta des m. chricothyroideus fest; Konsequenz aus diesem Befund sei, dass es unmöglich sei, von absoluten Aktivierungsmustern des m. cricothyroideus auf eine Beteiligung dieses Muskels auf vokalspezifische Grundfrequenz zu schließen, da die Aktivierungsmuster für die bewusste Grundfrequenzkontrolle sich von Vokal zu Vokal unterscheiden. Somit fehle die Vergleichbarkeit des Effekts eines bestimmten Ausmaßes an m. cricothyroideus-Aktivierung. Eine ausführlich Kritik zu Whalen et al. (1999) ist in Hoole (2006) zu finden. U. a. hält Hoole (2006) neben den hier teilweise angedeuteten Problemen mit der Methodik fest, dass diese Befunde ja nicht ausschlössen, dass – neben den mechanischen Gründen für intrinsische *f0* – Sprecher zusätzlich den m. cricothyroideus aktivieren, um intrinsische Grundfrequenz-Effekte aktiv zu verstärken; die Befunde in Whalen et al. (1999) könnten auch lediglich so gedeutet werden, dass es keine eins-zu-

eins-Relation zwischen m. cricothyroideus-Aktivität und vokalspezifischer Grundfrequenz gibt. Hoole, Honda, Murano, Fuchs und Pape (2004), Hoole (2006) als auch Hoole und Honda (2011) zeigen in [±*gespannt*]-Vokalpaaren des Deutschen eine durch Elektromyographie (EMG) ermittelte erhöhte Aktivität des m. cricothyroideus bei den ungespannten Vokalen bei einigen ihrer Versuchspersonen; dies lässt sich dahingehend deuten - und dies tun die Autoren - dass intrinsische Grundfrequenz aktiv eingesetzt werden *kann*, um Vokalkategorien wirksamer voneinander abzugrenzen. Dies lässt mechanische, passive Erklärungsmodelle für intrinsische Grundfrequenz unberührt, sondern erweitert die Erklärung zu einem hybriden Modell, in dem der Sprecher die Möglichkeit hat, zum Zwecke des *feature enhancement* (Diehl & Kluender, 1989; Kingston, 1992) zusätzlich den ohnehin vorhandenen, mechanisch erklärbaren Effekt der intrinsischen Grundfrequenz aktiv zu verstärken.

Hoole und Honda (2011) halten hierzu fest:

> As a final conclusion uniting both the consonant voicing and the vowel *If0* issues we suggest that it is a general feature of movement planning to take advantage of physical forces, where possible („go with the flow“). As a guideline for future work, it can be hypothesized that whenever an effect is assumed to be automatic and mechanical then it should be possible to demonstrate that it is fairly constant over speakers. On the other hand, the adoption of enhancement strategies will probably be more variable, reflecting the fact that speakers differ in clarity and, perhaps, in their sensitivity to acoustic differences.

Auch Berry und Moyle (2011) kommen zu einem vergleichbaren Schluss; sie messen die Kovariation von Intrinsischer Grundfrequenz, *voice onset time* (VOT) und *voiceless interval duration* (VID) in in einen Trägersatz eingebettete /CVd/-Äußerungen, mit C=/p,t,k,s/ und V=/i,u,a,æ/, während diese in habitueller und angehobener Tonlage gesprochen werden; sie stellen aber eine solche Kovariation von VOT und VID mit der intrinsischen Grundfrequenz (die für einen Automatismus sprechen würde, wie Holt, Lotto und Kluender (2001) vorgeschlagen hatten) nur bei männlichen Sprechern unter der habituellen Tonlage fest, nicht für angehobene Tonlagen und generell nicht für weibliche Sprecher, und schließen daher, Hoole und Honda (2011) zustimmend: „Thus while some variation in these acoustic measures may derive automatically from extralaryngeal influences, talkers *can* actively modify laryngeal behavior to enhance specific acoustic cues“(Berry & Moyle, 2011, Seite EL369, Hervorhebung von mir).

Wenn also die intrinsische Grundfrequenz *auch* zu *feature enhancement* genutzt wird, liegt nahe, anzunehmen, dass nicht *F1* alleine das beste akustische Vokalhöhenkorrelat ist, sondern dass *F1* zur Grundfrequenz des Sprechers in Relation gesetzt wird; in gewisser Weise muss man also davon ausgehen, dass *F1* durch die gegebene *f0* normalisiert wird. Genau hiervon gingen einige Ansätze der sogenannten *intrinsischen* Normalisierung aus, die wir im Folgenden etwas genauer betrachten wollen.

1.6 Der *F1-f0*-Abstand als ein Maß vokalintrinsischer Normalisierungsmethoden

Die Idee der rein *intrinsischen* Normalisierung, also der Normalisierung nur anhand eines gegebenen Formantmusters ohne weitere Betrachtung des Kontexts, geht davon aus, dass man, wenn man die Vokale mehrerer Sprecher in der durch *F2* auf der x-Achse und *F1* auf der y-Achse aufgespannten Ebene abbildet und hierbei Überlappungen feststellt, lediglich die falschen Parameter ausgesucht hat. Das Ziel müsse also sein, zu den mehr oder minder invariant wahrgenommenen Vokalklassentokens die passenden *invarianten* Parameter zu finden, die die sprecherbedingte Variabilität also auf ein Minimum reduzieren. Wir haben weiter oben darauf hingewiesen, dass höhere Formanten ein direkter cue für die Vokaltraktgröße ist (aber perzeptuell offenbar weniger wichtig), und dass die Grundfrequenz ein eher indirekter Cue für die Vokaltraktgröße ist, der aber offenbar relativ prominent wahrgenommen wird; desweiteren stellten wir fest, dass – gleiche Bedingungen vorausgesetzt – *f0* mit der Vokalhöhe variiert. Ein schon sehr früh entwickelter, da naheliegender Ansatz war es daher, dass die *Verhältnisse* der Formanten, oder überhaupt der spektralen Prominenzen (was dann auch die Grundfrequenz einschließen würde), eine Rolle für die Wahrnehmung der Klangfarbe der Vokale spielen sollte, so wie die Frequenz*verhältnisse* in musikalischen Akkorden die Klangfarbe des Akkords bestimmen, unabhängig von der Lage dieser Frequenzen. Schon Potter und Steinberg (1950, Seite 812) machten auf diese Analogie aufmerksam: „a certain spatial pattern of stimulation on the basilar membrane may be identified as a given sound regardless of position along the membrane. Musical chords, for example, are identified in this manner. Thus, the ear can identify a chord as a major triad, irrespective of its pitch position ... An implication is that final interpretation of audible form in the brain is by the same process as interpretation of visible form, a characteristic assumed in the Gestalt theory.“

Wenn man nun also davon ausgeht, dass es auf der Basilarmembran beim Hören eines Vokals zu einer Frequenz-Orts-Transformation kommt, wobei eine Vokalkategorie ein bestimmtes Muster der Anregung erzeugt, und dass dieses Muster erkannt wird, egal, *wo* auf der Basilarmembran dieses Muster aufscheint (so wie es für die Gestalterkennung irrelevant ist, wo ein Objekt auf der Retina des Auges abgebildet wird), muss man notwendigerweise mit einrechnen, dass die Basilarmembran Frequenzen nicht linear nach der Hertz-Skala repräsentiert, sondern in die sogenannten *kritischen Bänder* eingeteilt ist, die für niedrige Frequenzen schmaler als für höhere Frequenzbereiche sind (Fletcher, 1940). Dementsprechend erwies es sich als hilfreich, zu Normalisierungszwecken eine Skala zu verwenden, die dieser Einteilung in kritische Bänder mehr entspricht, wobei sich die verschiedenen Ansätze unterscheiden. Die oben erwähnten *Verhältnisse* spektraler Prominenzen zueinander lassen sich dann, wenn die Frequenzen dieser Prominenzen auf eine eher auditorisch linearen Skala transformiert wurden, auch als Abstände zwischen den Prominenzen beschreiben. Wie Johnson (2005) in seinem Vergleich verschiedener Ansätze zu intrinsischer Normalisierung hinweist, ergeben sich erstaunliche Ähnlichkeiten zwischen den Ansätzen, wenn man erstens bedenkt, dass die verwendeten Skalen starke Ähnlichkeiten aufweisen (so wie

die (verschiedenen Versionen der) BARK-Skala (Zwicker, 1982; Traunmüller, 1990a) dem ebenfalls gelegentlich verwendeten Logaritmus der Frequenzen (z.B. in Sussman (1986) oder (J. D. Miller, 1989)) nicht unähnlich sind (siehe zu dieser Frage auch die Diskussion in J. D. Miller (1989)), und dass die verwendeten Formeln mathematisch zueinander äquivalent sind (wegen $log(x) - log(y) = log(x/y)$; *Abstände* entsprechen her also *Verhältnissen*). Ein ausgeprägter Unterschied ist jedoch, ob die Grundfrequenz (in welcher Form auch immer) in den den Vokalraum definierenden Rahmen inkorporiert wird. So berechnete Peterson (1961) die logarithmisierten Frequenzen der Formanten 2 bis 4 in Relation zum entsprechenden Wert des ersten Formanten, und Sussman (1986) kommt ebenfalls ohne Grundfrequenz aus, indem er jeweils den Logarithmus des Verhältnisses der Formanten 1 bis 3 zum bereits in 1.3 erwähnten geometrischen Mittel der ersten drei Formanten, dass als Korrelat der Vokaltraktlänge gilt, verwendet.

Die Grundfrequenz findet Eingang in die Normalisierungsansätze von z.B. Syrdal und Gopal (1986) und J. D. Miller (1989) und Hirahara und Kato (1992). Syrdal und Gopal (1983) und Syrdal und Gopal (1986) bauen auf die bereits in 1.4 erwähnte spektrale Integration von nicht mehr als 3 bis 3.5 Bark auseinanderstehender Formanten (Chistovich & Lublinskaya, 1979, „center of gravity“) auf und kombinieren diesen Gedanken mit den oben erwähnten relativen Abständen zwischen Formanten als Normalisierungsmethode. Ihre berühmt gewordene 3-Bark-Hypothese lässt sich kurz so zusammenfassen, dass phonologische Merkmale sich, durch Transformation der Frequenzwerte in Bark und Ermittlung der Abstände zwischen den spektralen Prominenzen, ebenfalls binär beschreiben lassen als Erfüllung oder Nicht-Erfüllung der Bedingung, dass das in Frage stehende Prominenzpaar weniger als drei Bark voneinander entfernt ist und somit spektral integriert wird. Wie man an der Tabelle in Abbildung 1.3 sieht, umfasst die spektrale Integration nicht ausschließlich adjazente spektrale Prominenzen, sondern auch den Abstand zwischen *F4-F2*, der ebenfalls auf unter drei Bark fallen können soll.

Zu diesen Prominenzpaaren gehört auch der Abstand von *F1* und *f0*, transformiert auf die Bark-Skala; dieser Abstand repräsentiert das phonologische Merkmal $[\pm high]$; Der *F3-F2*-Abstand repräsentiert die Unterscheidung zwischen vorderen und hinteren Vokalen, usw.. Das schließt nicht aus, dass nicht generell *F1-f0* [Bark] das gegenüber *F1* bessere Vokalhöhenkorrelat ist, und *F3-F2* [Bark] nicht besser als *F2* alleine ist, was dadurch deutlich wird, dass in vielen Studien Abbildungen mit diesen beiden Dimensionen benutzt wurden, und schon auf diese Weise das Ausmaß an Variation gegenüber der *F2-F1*-Ebene erheblich reduziert werden konnte.

J. D. Miller (1989) repräsentiert Vokale in einem dreidimensionalem Raum: eine Dimension ist $log(\frac{F3}{F2})$, die zweite ist $log(\frac{F2}{F1})$, und die dritte ist $log(\frac{F1}{Sensorische\ Referenz})$, wobei die $Sensorische\ Referenz$ abgeleitet ist vom geometrischen Mittel der Grundfrequenz des Sprechers, gemessen über einen gewissen Zeitraum. Miller bietet hiermit u. a. sogar einen Ansatz, die suprasegmental bedingte Variation der Grundfrequenz mit in die Berechnung der $Sensorischen\ Referenz$ einfließen zu lassen, und insofern ist sein Ansatz nicht allein eine *intrinsische* Methode. Jedenfalls wird bei Millers Ansatz der vokalintrinsische Wert der Grundfrequenz weniger berücksichtigt als z.B. als in Syrdal und Gopal (1986), dafür aber die eher generelle Lage der *f0*.

TABLE III. Vowel classification based on critical distance features in five bark-difference dimensions.

	Dimensions				
Vowels	*F1–F0* <3 bark	*F2–F1* <3 bark	*F3–F2* <3 bark	*F4–F2* <3 bark	*F4–F3* <3 bark
/i/	+	−	+	+	+
/ɪ/	+	−	+	−	+
/ɛ/	−	−	+	−	+
/æ/	−	−	+	−	+
/ɜ/		−	+	−	−
/ʌ/	−	−	−	−	+
/a/	−	+	−	−	+
/ɔ/		+	−	−	+
/ʊ/	+	−	−	−	+
/u/	+	−	−	−	+

Abbildung 1.3: *Syrdal und Gobal (1986): 3-Bark-Hypothese. Tabelle mit binären Merkmalen, entnommen aus Syrdal und Gopal (1986, Seite 1091), mit originaler Abbildungsüberschrift.*

Hirahara und Kato (1992) schlagen vor, *F1-f0*[Bark] als Vokalhöhenkorrelat und *F2-f0*[Bark] als Zungenlagekorrelat zu verwenden, um sprecherbedingte Variation zu minimieren, also statt der klassischen Parameter *F1* und *F2* zur Grundfrequenz normalisierte Varianten dieser beiden Maße.

Wir wollen hier nicht ausführlicher beschreiben, wie erfolgreich Methoden, die versuchen, *intrinsische* Eigenschaften von Vokalen anzuwenden, um die Variation zwischen Sprechern z.B. im Rahmen von sozio-phonetischen zu minimieren, angewendet werden können, und wie sie im Vergleich mit den sogenannten *extrinsischen* Normalisierungsalgorithmen (z.B. Lobanov (1971); Nearey (1978)), die aus dem gegebenen Sprachmaterial eine statistische Beschreibung des Vokalraums erstellen und so gegebenen einzelne tokens in Relation zum gesamten produzierten Vokalraum setzen, abschneiden (siehe hierzu z.B. Disner (1980), Adank (2003) und Adank, Smits und van Hout (2004)). Es sei aber erwähnt, dass man sich die erwähnten Invarianzen nicht so vorstellen darf, dass nach der intrinsischen Normalisierung alle Sprechereigenheiten wegnormalisiert seien: so stellt Heid (1998, Kapitel 5.5) in einer Anwendung der Methode von Syrdal und Gopal (1986) auf einen größeren Korpus, der Aufnahmen von Männer- und Frauenstimmen enthielt, eine erhebliche Verringerung der Variabilität, die durch den Faktor Geschlecht zustandekommt, fest, aber auch, dass im Vergleich zweier Sprecher aus der selben Geschlechtsgruppe sich die Unterschiede eben nicht aufheben; nun kann dies sogar erwünscht sein, da dies die individuelle Variation

in der Artikulation widerspiegelt, was auch Ausdruck der Zugehörigkeit des Sprechers zu einer (z.B. sozio-phonetisch zu definierenden) Sprechergruppe sein kann, und gerade solche Information will man beibehalten; jedenfalls darf man aber grundsätzlich nicht die falsche Vorstellung haben, dass durch intrinsische Methoden durch die angewendete Tranformation in Verhältnis- bzw. Abstandsmaße für eine Vokalkategorie ein Paar oder Triple oder Quadrupel usw. von wirklich invarianten Werten entsteht. Auch in der originalen Quelle (Syrdal & Gopal, 1986) wurde schon eine nicht unerhebliche Zwischen-Sprecher- und auch Inner-Sprecher-Variation der normalisierten Werte beschrieben.

Zusammenfassend lässt sich also sagen, dass durch intrinsische Methoden versucht wird, die gröbsten Variabilitäten im akustischen Output zwischen Geschlechts- und Altersgruppen (zwischen Kindern und Erwachsenen) zu minimieren, indem anhand des gegebenen Formantpatterns eines Vokaltokens jene Wertkombinationen gesucht werden, die vergleichsweise wenig variieren. Doch gibt es starke Hinweise darauf, dass dies bei weitem nicht ausreicht, um die Normalisierungsfähigkeiten menschlicher Hörer nachzubilden. Wie zahlreiche Experimente gezeigt haben, sind Erkennungsraten auch davon beeinflusst, dass der gegebene Laut in den vom gegenwärtigen Sprecher erzeugten „Rahmen“ passt, den man als Referenzrahmen bezeichnen könnte, und der über den Einzellaut hinausgeht; damit hat man also ein Koordinatensystem zu vermuten, in dem man eine relative Zuordnung vornehmen kann. So zeigen Ladefoged und Broadbent (1957), indem sie in einem synthetisch erzeugten Trägersatz generell die Formantfrequenzen änderten (*F1*, *F2*, oder beide kombiniert), dass in diesen Satz eingebettete Testwörter mit gleichbleibenden Formantfrequenzen anders wahrgenommen werden, obschon in diesen ja *intrinsisch* alles unverändert bleibt. Ähnliches wurde auch bei Ainsworth (1975) und Nearey (1989) festgestellt, und auch Gerstman (1968) schlägt eine solche Kalibrierung anhand eines Referenzrahmens vor. Assmann, Nearey und Hogan (1982) fanden höhere Erkennungsraten, wenn die Hörer ihrer Experimente mit Signalen desselben Sprechers konfrontiert werden, anstelle von Sprache verschiedener Sprecher. Man kann also davon ausgehen, dass der Referenzrahmen, den der einzelne Sprecher in einer über den Einzellaut hinausgehenden Äußerung aufspannt, die Erkennung und Zuordnung zu Kategorien beeinflusst, aber auch erleichtert, und zwar unabhängig von *intrinsischen* Eigenschaften einzelner Phone, insbesondere von Vokalen. Dennoch ist eine gewisse Involvierung der beschriebenen *intrinsischen* Eigenschaften bei der von menschlichen Hörern vorgenommenen Erkennung nicht zu leugnen.

1.7 *F1* und *f0* bei Traunmüller

Bislang haben wir die Arbeiten Traunmüllers zum Zusammenhang zwischen *F1* und *f0* noch nicht erwähnt, obschon wir darauf hingewiesen haben, dass sie sehr wichtig für die Frage sind. Dies hat damit zu tun, dass sich Traunmüller mit Zusammenhängen zwischen beiden Maßen nicht allein aus Gründen einer Normalisierung bezüglich der Vokalhöhe beschäftigt hat, weshalb wir seine Arbeiten gesondert besprechen wollen.

Sein wohl meistzitiertes Werk (Traunmüller, 1981) untersucht den Einfluss von synthetischen Ein-Formant-(*F'*)-Stimuli, bei denen systematisch die Grundfrequenz und der

Formant variiert wurde (und zwar über die gesamte Bandbreite von Frequenzen, die in natürlicher Sprache vorkommt, nämlich von 50 bis 700 Hz für *f0* und von 150 bis 1480 Hz für *F'*), auf die Vokalhöhenperzeption von Niederösterreichischen Hörern (die somit Sprecher eines ostmittelbairischen Dialektes waren, in dem 5 Vokalhöhendistinktionen vorherrschen), sowie den Einfluss der gleichen Manipulation in Kopiesynthesen natürlich gesprochener vorderer ungerundeter und gerundeter Vokale (Sprecher war Traunmüller selbst). Die Variation von *f0* und *F'* (bzw. *F1*) wurde in Bark-Schritten vorgenommen.

Abbildung 1.4 zeigt die perzeptiven Grenzen zwischen den 5 Vokalhöhen auf einer Ebene, die aufgespannt wird durch die Grundfrequenz auf der x-Achse und dem Bark-Abstand zwischen dem Formanten und *f0* auf der y-Achse. Über relativ weite Bereiche der *f0* und für die meisten Vokalhöhendistinktionen verlaufen die Grenzen relativ parallel zur x-Achse, was bedeutet, dass – zumindest innerhalb gewisser Grenzen – der *F'-f0*-Abstand vergleichsweise invariant verläuft, und die Vokalhöhendistinktion besser abbildet als *F'* alleine (was aber z.B. im *f0*-Bereich oberhalb von 350 Hertz für die Distinktion zwischen den Vokalhöhen 4 und 5 der Fall ist). Auch für die Kopiesynthesen bestätigte sich, dass der *F1-f0*-Abstand das gegenüber *F1* bessere Vokalhöhenkorrelat ist. Dies bedeutet also, dass, wenn man *f0* und *F1* gemeinsam erhöht, und hierbei den Bark-Abstand gleich lässt, sich die Vokalkategorisierung nicht ändert; Traunmüller vermutete hierbei einen tonotopischen Gestaltereknnungsprozess („tonotopic distance hypothesis“).

Diese starke Sichtweise der Dinge ist im Laufe der Zeit modifiziert worden, teilweise auch von Traunmüller selbst. Traunmüller (1984) beschreibt, wie auch schon Traunmüller (1981) und Traunmüller (1983), die Vokalerkennung als Prozess der tonotopischen Gestalterkennung, und behauptet eine vom Sprecher unabhängige Invarianz für die Formantabstände bei phonetisch gleichen Vokalen, solange diese Abstände 6 Bark nicht übersteigen; bezüglich des Vokalhöhenkorrelats *F1-f0*[Bark] stellt Traunmüller aber fest, dass Frauen für die gleichen Vokale offenbar einen kleineren Wert für diesen Abstand aufweisen als Männer und Kinder. Auch Traunmüller (1988) beschreibt solche Unterschiede zwischen Frauen auf der einen und Männern und Kindern auf der anderen Seite - auch unter Benutzung offenbar invarianterer Cues (wie *F1-f0*[Bark]) ist der durchschnittliche Vokalraum von Frauen peripherer.

Traunmüller (1985) beschreibt, dass die phonetische Identität auch unbeeinflusst bleibt, wenn man alle Formanten Bark-skaliert mit *f0* variiert, wobei diese Erhöhung der Werte als Abnahme der Sprechergröße wahrgenommen werde - z.B. als Wechsel von erwachsenem zu kindlichem Sprecher. Die von Fant (1975) festgestellte Nicht-Uniformität verschwindet also bei Anwendung der Bark-Skala. Bei gleichzeitiger paralleler Variation von *f0* und *F1* bleibt zwar die Vokalhöhe erhalten, es ändert sich aber die Perzeption des *vocal efforts* - von eher ruhiger Sprechweise in tiefen Lagen zu Rufen in hohen Lagen. Wie Traunmüller (1988) und Traunmüller (1991b) beschreiben, variiert auch *F2* in hinteren Vokalen mit *vocal effort*; *F3* und höhere Formanten bleiben allerdings unverändert. Diese Anhebung von Grundfrequenz, erstem und teilweise zweitem Formanten (wenn dieser tief ist), und damit die Annäherung an die eher die Ansatzrohrgröße definierenden höheren Formanten nennt Traunmüller *elevation*. Es existiert also eine durch Variation des *vocal efforts* bedingte Kovariation von *F1* und *f0* innerhalb des selben Sprechers (vgl. auch Traunmüller und

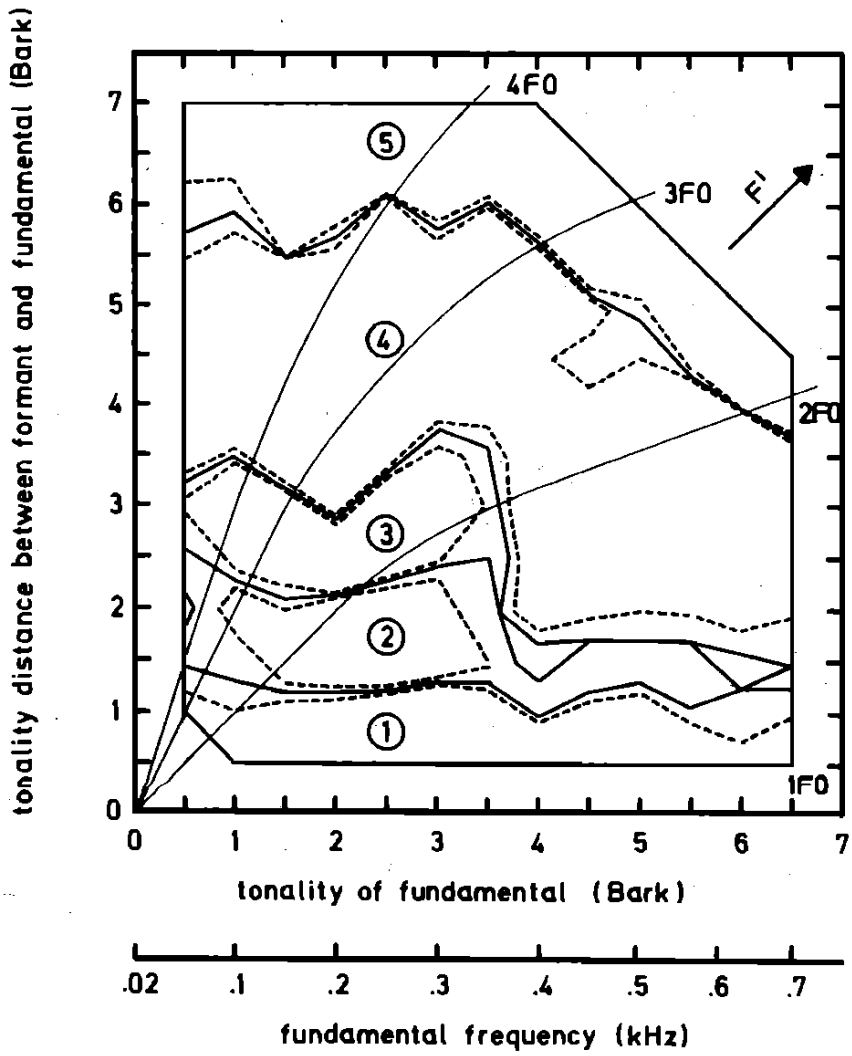

FIG. 4. One-formant stimuli: Dominantly perceived degrees of openness (thick lines) and regions of at least 50% agreement among subjects (dashed lines), shown together with positions of first 4 partials (thin lines). Encircled figures designate degree of openness (cf. Table II and Table III).

Abbildung 1.4: *Ergebnisse des Ein-Formant-Experiments aus Traunmüller (1981, Seite 1468), mit originaler Abbildungsüberschrift.*

Eriksson (2000), woraus auch Abbildung 1.5 stammt, Eriksson und Traunmüller (2002), sowie die Diskussion in Kapitel 2.2.1).

Die Ergebnisse aus Traunmüller und Lacerda (1987) (siehe auch die Diskussionen hierzu in Traunmüller (1991a, 1991b)) lassen vermuten, dass die *F1-f0*-Distanz ein umso wichtigeres Vokalhöhenkorrelat ist, je mehr Vokalhöhendistinktionen eine Sprache aufweist -

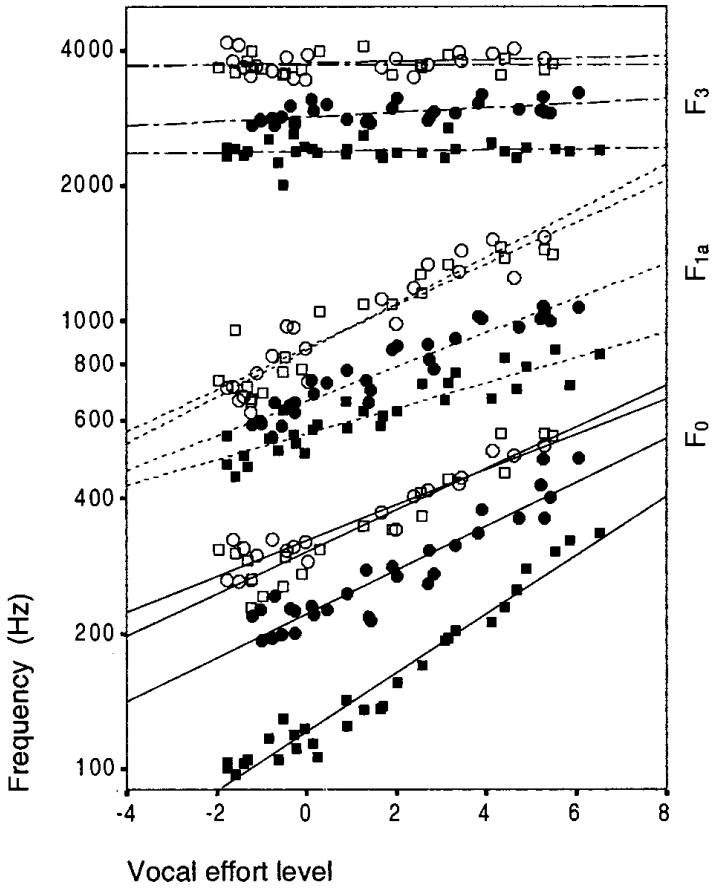

FIG. 5. Mean values of F_0, F_{1a}, and F_3, shown as a function of VEL for men (■), women (●), boys (□), and girls (○). Regression lines fitted to each variable (solid, dotted, broken lines) and speaker group. See also Table VI.

Abbildung 1.5: *Kovariation von F1 und f0 bei vocal effort-Variation, aus Traunmüller und Eriksson (2000, Seite 3446), mit originaler Abbildungsüberschrift.*

unter Sprechern des vokalhöhenarmen Türkischen gibt es weniger Versuchspersonen, die den *F1-f0*-Cue nutzen als unter Sprechern des vokalhöhenreicheren Schwedischen (und unter diesen – vermutet Traunmüller – mehr Personen, die den Cue nicht nutzen, als unter den ursprünglich von ihm (Traunmüller, 1981) untersuchten Niederösterreichern; vergleiche hierzu auch die oben beschriebene Diskussion um den aktiven Einsatz der Intrinsischen Grundfrequenz in vokalhöhenreichen Sprachen). In Traunmüller (1991b, Seite 127) schreibt Traunmüller hierzu: „the larger the number of distinctive degrees of openness in a subject's native language, the larger the probability that he will behave in accordance with the tonotopic distance hypothesis".

Hoemeke und Diehl (1994) finden für amerikanisches Englisch, dass der *F1-f0*-Abstand

ein guter Cue für die Distinktion zwischen /ɪ/ und /ɛ/ ist, *nicht* aber für die Unterscheidung zwischen /i-ɪ/ oder zwischen /ɛ-æ/. Dies spricht dafür, dass die These von Syrdal und Gopal (1986), dass wegen der spektralen Integration innerhalb eines Abstandes von 3 Bark der *F1-f0* Abstand *hauptsächlich* für die [$\pm high$]-Distinktionen geeignet sei, zumindest für Amerikanischen Englisch richtig ist. Die fünf-stufige Vokalhöhendistinktion im ostmittelbairischen Dialekt, den Traunmüller (1981) untersuchte, sei im Gegensatz zu den hier verwendeten Vokalpaaren durch eine *echte* Vokalhöhendistinktion im phonologischen Sinn bestimmt, während zumindest das /i-ɪ/-Paar des Amerikanischen Englisch, bei dem der *F1-f0*-Cue versagt, eher durch [$\pm tense$] unterschieden sei, so Hoemeke und Diehl (1994).

In einer Untersuchung hinterer Vokale des Amerikanischen Englisch fanden Fahey, Diehl und Traunmüller (1996), dass *F1-f0* ein besserer Cue als *F1* ist für die Distinktion zwischen /u/-/ʊ/ und /ʊ/-/ʌ/, nicht aber für die Paare /ʊ/-/ɔ/, /ɔ/-/ɑ/ und /ʌ/-/ɑ/. In ihrer Diskussion argumentieren sie für eine mögliche Rolle des Abstandmaßes *F2-F1*, und argumentieren generell eher gegen die spektrale Integration innerhalb eines Bereichs von 3 Bark.

Für das Französische fanden Ménard, Schwartz, Boë, Kandel und Vallée (2002) eine hohe Wichtigkeit des *F1-f0*-Abstandes in Bark für die Vokalhöhenunterscheidung, aber keine Hinweise darauf, dass in dieser Sprache die Vokalhöhendistinktion über *F1-f0*[Bark] hauptsächlich binär (über [$\pm < 3Bark$]) organisiert wäre. Stattdessen gilt einfach, dass der Vokal umso offener perzibiert wird, je größer der *F1-f0*-Abstand ist.

Eine abnehmende Wichtigkeit des *F1-f0*-Cues für Grundfrequenzen $< 150Hz$ (die aber auch Traunmüller (1983) selbst schon beschrieben hatte und in Traunmüller (1990b) bestätigte) sowie eine Abnahme der Wichtigkeit des Abstandmaßes zwischen erstem Formant und Grundfrequenz mit zunehmender Vokalhöhe stellte Di Benedetto (1987, 1994) über akustische Analysen und mehrere Perzeptionsexperimente fest. Di Benedetto (1994) stellte die These auf, dass der *F1-f0*-Abstand nur dann wichtig ist, wenn beide Parameter ausreichend hoch sind. Für tiefere Werte von *F1* (in hohen Vokalen) und für *f0*-Werte unter 150 Hz vermutete sie eine Bewertung des ersten Formanten gegenüber einem anderen Ankerpunkt als der *f0*, nämlich dem Ende der Skala (sprich, der *F1*-Wert an sich wird dann bewertet). In Di Benedetto (2003) fand sie in einer akustischen Analyse eines Korpus' (der aus Sprachmaterial von je 2 Männern und Frauen bestand) keine Vorteile des *F1-f0*-Abstandes gegenüber *F1* zur Vokalhöhendistinktion (siehe hierzu auch Hillenbrand und Gayvert (1993)), noch generell zur Normalisierung der sprecher- und geschlechtsspezifischen Unterschiede. Dies ist allerdings erklärbar, wenn man die sich oben bereits erwähnten Resultate von Heid (1998) vor Augen hält, der in größeren Sprechergruppen eine starke Reduktion der geschlechtsbedingten Variation feststellte, nicht jedoch im Vergleich von Sprecherpaaren (des selben Geschlechts) – möglicherweise sind also die Ergebnisse in Di Benedetto (2003) durch die starke individuelle Variation beeinflusst, und die Prominenzabstandsmaße (wie in Syrdal und Gopal (1986)) normalisieren nur in größeren Gruppenstärken.

Fahey und Diehl (1996) fanden, dass kaum Unterschiede in der perzeptuelle Nutzung des *F1-f0*-Abstandes bestehen, auch wenn man die niederen Harmonischen (also auch die Grundfrequenz an sich) herausfiltert - die Grundfrequenzinformation, die durch die Abstän-

de der Harmonischen ja noch nutzbar bleibt, reicht also aus, was gegen die ursprüngliche Traunmüller'sche Hypothese spricht, dass die Vokalhöhenperzeption durch die Anregung der Basilarmembran durch die erste Harmonische und den ersten Formanten bestimmt wird.

In Traunmüller (1991a) untersuchte Traunmüller die Kontextabhängigkeit des *F1-f0*-Cues, und fand, dass manche Hörer sich offenbar eher an dem gegenwärtigen *f0*-Wert orientieren, und andere eher an einer prosodischen *baseline*, die man sich in etwa als lineare Verbindung der *f0*-Minima in einer Äußerung vorstellen müsse (wobei diese laut Traunmüller und Eriksson (1995a, 1995b) auch bei Variation der Spannweite der *f0* für jeden Sprecher invariant bleibt).

In Traunmüller (1991b) schlägt der Autor vor, dass, da dieser Basis-Wert vom Hörer geschätzt werden müsse, er dies auch für geflüsterte Sprache tun könne - für geflüsterte Sprache finden Eklund und Traunmüller (1997) aber durchaus geringere korrekte Klassifizierungen als für phonierte Sprache. Coleman, Grabe und Braun (2002) gehen der Frage nach, ob die Wahrnehmbarkeit und Unterscheidbarkeit von Intonationskonturen und distinktiven Tönen bei geflüsterter Sprache, bei der also die Stimmlippen nicht schwingen, sondern stattdessen das sogenannte Flüsterdreieck geöffnet ist, und somit keine Grundfrequenz vorhanden ist, damit zu erklären sei, dass die Kehlkopfhöhe so variiere wie in modaler Stimmgebung. Sie finden eine zwar schwächer ausgeprägte, aber eindeutig vorhandene Kehlkopfhöhenvariation, wie man sie auch bei modaler Stimmgebung erwarten würde, und schließen daraus, dass spektrale Unterschiede, die durch die variable vertikale Kehlkopflage erklärbar sind, dem Hörer als akustischer Cue für die Perzeption von pitch- und Ton-Variation bei geflüsterter Sprache dienen könne. So gesehen ist also vorstellbar, dass die prosodische *baseline*, die Traunmüller meint, auch ohne vorhandene Grundfrequenz wahrnehmbar ist. Doch auch dieser Umstand spricht dagegen, dass der Effekt des *F1-f0*-Cues durch eine Anregung eines bestimmten Musters auf der Basilarmembran zu erklären ist.

Schon in Traunmüller (1990b, Seite 2018, Fußnote 4) schreibt Traunmüller, er betrachte den *F1-f0*-Abstand in Bark „ just as an empirical description of where *F1* is to be expected as a function of *f0*". Zusammenfassend lässt sich also sagen, dass sich der *F1-f0*-Abstand in Bark als weniger invariantes Merkmal für Vokalhöhendistinktionen herausgestellt hat, als zunächst zu vermuten war; seine Wichtigkeit variiert offenbar mit der Anzahl der Vokalhöhendistinktionen in der untersuchten Sprache, aber auch unter Hörern einer Sprache gibt es beachtliche Unterschiede in der Nutzung des Cues (ein Umstand, der stark an das oben gemachte Zitat aus Hoole und Honda (2011) erinnert); teilweise beziehen sich diese Unterschiede auch darauf, ob der Hörer *F1* zu einem momentanen Wert von *f0* oder eher zu einer prosodischen *baseline* relativiert. Generell variiert die Wichtigkeit des *F1-f0*-Cues auch mit den jeweiligen Lagen der Grundfrequenz und des ersten Formanten. Generell kovariieren *f0* und Formanten allgemein mit Geschlecht und Alter (wobei hier Traunmüller die Unterscheidung zwischen Männern, Frauen, und Kindern meint, deren Werte laut Traunmüller (1985) in den Parametern *f0* und den höheren Formanten ab *F3* je circa ein Bark auseinander liegen, was auf eine Spezialisierung des menschlichen Gehörs auf Sprachwahrnehmung schließen lässt). Eine Kovariation von *f0* und *F1* (und auch von *F2* bei hinteren Vokalen)

lässt sich für Variationen von *vocal effort* feststellen, während hierbei die Formanten ab *F3* unverändert bleiben (was Normalisierungen, die z.B. *F3-F2*[Bark] mit einschließen, für Daten mit *vocal effort*-Variation stark einschränkt).

1.8 Kontext und Beeinflussung der Vokalhöhe durch *f0*

Ein Einwand gegen die These, dass die Formanten oder zumindest der erste Formant zur Grundfrequenz normalisiert werden würden, ist der z.B. von Whalen und Levitt (1995) vorgetragene, dass in der alltäglichen Kommunikation *f0* hauptsächlich auch die Funktion der Intonation erfüllt, und deshalb ein pitch-Akzent-tragender Silbenkern bezüglich der Vokalhöhe anders wahrgenommen werden müsste, d.h. dass beide Funktionen (so *f0* denn eine vokalhöhendefinierende Funktion hat) einander im Wege stehen würden. Hierzu ist erstens zu sagen, dass manche Modelle, die den Abstand zwischen erstem Formanten und Grundfrequenz als vokalhöhendefinierendes Merkmal sehen, *nicht* den vokalintrinsischen Wert als Ankerpunkt annehmen, sondern eine Art Mittelwert (wie in J. D. Miller (1989)) oder eine Art Basis-Wert (wie in Traunmüller (1991a)) der Grundfrequenz, der über einen die Dauer eines Phons hinausgehenden Zeitraum bestimmt wird. Dies reduziert erheblich die Variabilität des Ankerpunkts *f0* (wobei der Basis-Wert Traunmüllers in Traunmüller und Eriksson (1995a) als für jeden Sprecher spezifischer Wert, der sich aber bei expressiver Variation *nicht* ändert, also invariant bleibt, beschrieben wird, und somit unsere gemachte Annahme noch mehr unterstützt als Millers Mittelwert).

Andererseits stellten Syrdal und Steele (1985) fest, dass eine dynamische Änderung der Grundfrequenz mit einer Änderung des ersten Formanten positiv korreliert, wobei die Invarianz des *F1-f0*-Abstandes nicht erhalten bleibt, die von Syrdal und Gopal (1986) vorgeschlagene Separierung in hohe und nicht-hohe Vokale durch das 3-Bark-Kriterium allerdings schon. Diese Kovariation der Grundfrequenz und des ersten Formanten ist verwandt mit der Variation mit *vocal effort*, die Traunmüller und Eriksson (2000) beschrieben haben, und liegt auch nahe, wenn man bedenkt, dass zur Prominenzmarkierung nicht allein *f0* in Form eines *pitch accents*, sondern auch die Intensität eingesetzt wird (Kochanski et al., 2005), was zumeist mit einer Erhöhung des Öffnungsgrades (und damit einer Erhöhung des ersten Formanten) einhergeht[4].

Außerdem wird – bezüglich der Intrinsischen Grundfrequenz, deren Funktion als *feature enhancer* Whalen und Levitt (1995) in ihrer oben genannten Kritik bezweifeln – genau unterschieden zwischen der Intrinsischen Grundfrequenz und dem sogenannten *intrinsic pitch*, also der Wahrnehmung der Höhe des Grundtons - man stellt hierbei nämlich fest, dass für die vokalintrinsische Grundfrequenzvariation kompensiert wird, so dass, wenn beispielsweise /i/ und /a/-Tokens mit identischer Grundfrequenz präsentiert werden, und der Grundton bewertet werden soll, dieser bei /i/ als tiefer bewertet wird als bei /a/ (vgl. z.B. Chuang

[4]Für einen Vergleich der Effekte der linguistischen Prominenzmarkierung und des *vocal efforts* siehe Mooshammer (2010)

und Wang (1978), Stoll (1984), Fowler und Brown (1997), oder Niebuhr (2004)). Interessanterweise zeigen Pape und Mooshammer (2008), dass diese Effekte bei Sprechern/Hörern einer Sprache mit vielfältiger Vokalhöhendistinktion stärker ausgeprägt sind als bei Sprechern/Hörern von Sprachen mit nur wenigen Vokalhöhendistinktionen, was ein weiterer Beleg dafür ist, dass Sprecher/Hörer vokalhöhenreicher Sprachen die gegebene *f0* auch als *intrinsisches*, vokalhöhendefinierendes Merkmal wahrnehmen, während Sprecher/Hörer vokalhöhenarmer Sprache dies weniger oder nicht tun. Reinholt Petersen (1986) beschreibt denn auch die zwei Gesichter der vokalintrinsischen Grundfrequenzvariation[5] so: auf suprasegmentaler Ebene übt sie einen den Intonationsverlauf „störenden" Einfluss aus, für den kompensiert werden muss (->*intrinsic pitch*), auf segmentaler Ebene hingegen ist sie *gleichzeitig* ein perzeptiver Cue für die Phonemidentität, zumindest in manchen Sprachen (vgl. auch die Arbeit von Fowler und Brown (1997), auf die wir an geeigneter Stelle genauer eingehen werden).

Eine wichtige Quelle der Variation innerhalb des selben Sprechers haben wir bisher noch nicht einmal erwähnt, da Betrachtungen hierzu den Rahmen dieser Arbeit sprengen: die Variation, die durch Koartikulation, also der Überlappung artikulatorischer Gesten, entsteht. Sie sorgt nicht nur dafür, dass die Formanten der Vokale von Sprechern eben nicht *steady-state* sind, sondern über weite Teile der Dauer des Vokals als Transitionen auftreten, sondern auch dafür, dass beispielsweise in einer V_1CV_2-Folge sich die Vokale gegenseitig beeinflussen. Für uns ist an dieser Stelle interessant, dass Reinholt Petersen (1980) Evidenz dafür bietet, dass die intrinsischen Grundfrequenzen der Vokale in V_1CV_2-Folgen durch Koartikulation beeinflusst sich aneinander anpassen; dieser Koartikulationseffekt auf die Grundfrequenz scheint sogar deutlicher ausgeprägt als jener auf die Formanten. Denkt man sich nun diese beiden Änderungen – die in der intrinsischen Grundfrequenz und jene in den Formanten – auf einer adäquateren Skala als der benutzte Hertz-Skala präsentiert, so könnte man spekulieren, ob durch diese Kovariation durch Koartikulation der akustische Output nicht so verändert wird, dass das auditiver Ergebnis vielleicht weniger variant von Kontextbeeinflussung gehört wird, als man bislang dachte.

1.9 Formanttuning

Klassisch trainierte Sängerinnen, insbesondere Sopranistinnen, weisen im höheren Bereich ihrer *Tessitura*, also ihres Phonationsumfangs, Grundfrequenzen auf, die über der für das jeweilige Individuum gewöhnlichen Lage des ersten Formanten liegen können; außerdem kann es vorkommen, dass die Harmonischen des Quellsignals, also die Grundfrequenz und ihre Obertöne, zwischen den Resonanzfrequenzen des Vokaltrakts liegen, die dann eben nicht angeregt werden. Außerdem weisen hohe Frauenstimmen nicht den sogenannten Sängerformanten, also eine Clusterung der Formanten 3 bis 5 (siehe hierzu u.a. Sundberg (1987), aber auch Detweiler (1994)), auf, der zumindest in klassisch ausgebildete Männerstimmen vorzufinden ist und diesen ermöglicht, auch über ein stark besetztes Symphonieorchester

[5] Vergleichbares gilt aber natürlich auch für die Einflüsse auf die Grundfrequenzvariation durch benachbarte $[\pm stimmhaft]$ -Obstruenten

hinweg hörbar zu sein. Um all diesen ungünstigen Voraussetzungen entgegenzuwirken, *tunen* Sopranistinnen den ersten Formanten in die Region, in der die Grundfrequenz sich befindet, um diese Resonanz überhaupt zu nutzen, und um hierdurch eine ausreichend intensive Stimmgebung zu erreichen (Bresch & Narayanan, 2010; Carlsson & Sundberg, 1992; Garnier, Henrich, Smith & Wolfe, 2010; Joliveau, Smith & Wolfe, 2004a, 2004b; Lindblom & Sundberg, 1971; Sundberg, 1975, 1977, 1987); dies ermöglicht es auch ihnen, über begleitende Instrumente hinweg hörbar zu sein. Henrich, Smith und Wolfe (2011) zeigten kürzlich, dass nicht nur Sopranistinnen in bestimmten Lagen[6] (ihres Grundfrequenzbereichs) dieses Tuning verwenden: Altistinnen *tunen* die erste Resonanz ihres Vokaltraktes in hohen Lagen auch zur Grundfrequenz, in tiefer Lage jedoch zur zweiten Harmonischen (also zu $2 * f0$); letzteres tun auch Tenöre, während Bässe die erste Resonanzfrequenz in den Bereich des dreifachen ihrer Grundfrequenz, also zur dritten Harmonischen (zum zweiten Oberton der Grundfrequenz) hin, *tunen*. Weitere Hinweise zum Formanttuning auch bei männlichen Sängern sind in D. Miller und Schutte (1994) zu finden.

Für den bekanntesten dieser Effekte, das Formanttuning der Sopranistinnen, konnte gezeigt werden, dass sie, um den ersten Formanten in den Bereich der Grundfrequenz zu *tunen*, den Öffnungsgrad des Kiefers variieren, wie Bresch und Narayanan (2010) eindeutig anhand von Magnetresonanztomographiedaten darlegen.

Titze (2004) beschäftigte sich mit *resonanten* Sprech- und Singstimmen, wie sie professionelle Stimmbenutzer wie z.B. Bühnenschauspieler oder Opernsänger aufweisen, und modellierte hierzu die akustische Quelle-Filter-Interaktion, die offenbar hauptsächlich über Änderungen in der epipharyngalen Röhre gesteuert wird, die hierbei durch Anpassung des Verhältnisses der akustischen Impedanz des Ansatzrohrs und der der Glottis das Schwingen der Stimmlippen erleichtern. Titzes Ziel war, die Adjustierung zu finden, die gleichzeitig die Wandlung aerodynamischer Energie an der Glottis in akustische Energie optimiert und den Transfer dieser akustischen Energie maximiert. Hierbei spielt der erste Formant eine wichtige Rolle, da die beiden genannten Eigenschaften nach Titzes Ergebnissen zusammengenommen dann ein optimales Ergebnis liefern, wenn die Grundfrequenz, oder die zweite oder dritte Harmonische (also der erste bzw. zweite Oberton der Grundfrequenz) knapp unterhalb der Mittenfrequenz des ersten Formanten liegen. Ein exaktes Formanttuning, wie es in den meisten der oben genannten Beiträge angenommen wurde, ist laut Titze (2004) nicht optimal, da die Energiewandlung an der Glottis bei exakter Übereinstimmung von $k * f0$ (mit $k = 0, 1, 2$) und *F1* nicht mehr günstig ist. Jedenfalls zeigen auch diese Erkenntnisse, dass durch eine Anpassung des ersten Formanten an die Grundfrequenz oder ihre ganzzahligen Vielfache (wenn auch nicht exakt an diese angepasst, sondern leicht darüber) der akustische Output, der von den Lippen abstrahlt, stark ansteigen kann[7], und zwar um bis zu 10 dB.

Wir wollen hier nicht behaupten, dass „gewöhnliche“ Sprecher oder Amateursänger

[6]Diesem Formanttuning sind natürliche Grenzen gesetzt, da der Kiefer nicht beliebig weit geöffnet werden kann und sehr hohe Grundfrequenzen nun unabänderlich *über* der ersten Resonanz liegen; in höheren Sopranlagen sind daher Vokale nicht mehr unterscheidbar.

[7]Auch hier gilt natürlich der Energieerhaltungssatz, wie Titze feststellt; die Effekte beruhen darauf, dass bei der Energiewandlung Verluste an andere Energieformen als an die akustische minimiert werden.

von der genannten Strategie des Formanttunings Gebrauch machen würden; dennoch ist es erstaunlich, dass die angewendeten Methoden der *Stimmbildung*, die teilweise über Generation von Pädagogen hinweg mehr oder minder über Versuch und Irrtum entwickelt wurden, das (auch von den Pädagogen selbst unbewußt angesteuerte) Ziel haben, die Phonation durch eine Anpassung des ersten Formanten an die Grundfrequenz[8] zu erleichtern und dadurch die stimmlichen Ausdrucksmöglichkeiten zu optimieren. Es darf hierzu nicht unerwähnt bleiben, dass die Sprecher im ersten experimenten Kapitel dieser Dissertation professionelle Stimmbenutzer sind, und deshalb nicht ausgeschlossen werden kann, das auch sie in vermutlich erhaltenem Stimmtraining vergleichbare Strategien erlernt haben könnten.

Ein weiterer Zusammenhang zwischen dieser Art der Interaktion von Quelle und Filter mit der Sprache „gewöhnlicher“ Sprecher, die kein Stimmtraining erhalten haben, liegt möglicherweise in der *undersampling hypothesis*, also der Hypothese, dass die größeren Vokalräume bei weiblichen Sprechern nicht nur mit einer sozio-phonetisch regulierten „deutlicheren“ Aussprache, oder mit der Nicht-Uniformität der Unterschiede zwischen weiblichen und männlichen Ansatzrohren zu tun haben, sondern vielmehr mit dem Umstand, dass die höhere Grundfrequenz weiblicher Sprecher, die naturgemäß größere Abstände im Obertonspektrum hervorruft, dazu führt, dass die Resonanzen des Vokaltraktes kompensatorisch angepasst werden müssen, um diese überhaupt anzuregen, um Formanten ausreichender Amplitude zu erzeugen (Ryalls & Lieberman, 1982; Diehl, Lindblom, Hoemeke & Fahey, 1996; Simpson, 2001; Chládková, Boersma & Podlipský, 2009). In den genannten Arbeiten wird umgekehrt aber auch darauf hingewiesen, dass die geräuschhaftere Anregung durch Frauen, also der Einsatz von *breathy voice*, den man zumindest bei manchen Sprecherinnen einiger Varietäten des Englischen nachweisen kann, eine laryngale Strategie zur Kompensation des spektralen *undersamplings* sein könnte.

Maurer und Kollegen stellen die Behauptung auf, dass, wenn Kinder, Frauen und Männer Vokale bei der gleichen Grundfrequenz produzieren (einer Aufgabe, die dem Singen doch sehr nahe kommt), die Formantunterschiede im Bereich unter einer bestimmten Frequenz – die je nach Studie zwischen 1500 Hz und 2500 Hz variiert – zwischen diesen Sprechergruppen sich mehr oder weniger aufheben würden (Maurer & Landis, 1995, 1996) (vergleichbares für gesungene Vokale in Bloothooft und Plomp (1985)), oder doch zumindest keinen Geschlechtsunterschied in *F1* und *F2* bei [uː,oː,aː] und bei *F1* von [eː,iː], sowie keine Unterschiede in *F1* zwischen Erwachsenen und Kindern, sowie – unabhängig von Alter – nahezu identische Werte für die ersten beiden Formanten für [uː] (Maurer, Cook, Landis & d'Heureuse, 1991) (siehe auch (Maurer, Hess & Gross, 1996; Maurer & Klinkert, 1997)), was zu der starken Schlussfolgerung in Maurer, D'heureuse und Landis (2000, Seite 73) führt:

> . . . it is doubtful, at least for the lower formants, that they relate to vocal tract sizes. Thus, the relationship between vocal tract size and formant patterns in

[8] ... und durch andere Mittel, wie z.B. durch die Herausbildung des sogenannten *Sänger-* bzw. *Sprecherformanten*, also einer sehr breitbandigen Anhebung der Frequenzen um 3 kHz; dies ist ein Bereich, in dem z.B. Orchesterinstrumente nur vergleichsweise wenig Energie aufweisen.

general is more complex than is usually assumed...

Wir wollen dieser starken Behauptung nicht folgen, und versuchen, Maurers Befunde zu erklären: Es ist bekannt, dass die ausgeprägtesten Unterschiede zwischen den Ansatzrohren von Frauen und Männern die durch die zweite Kehlkopfabsenkung während der Mutation bedingte relative Vergrößerung des pharyngalen Raums bei Männern ist (Goldstein (1980), Fitch und Giedd (1999), vergleiche auch Ménard, Schwartz, Boë und Aubin (2007), d.h. nicht nur ist bei Erwachsenen eine Vergrößerung des Ansatzrohrs gegenüber Kindern festzustellen, sondern ein nicht-uniformes Wachstum, besonders ausgeprägt bei Männern, so dass sich auch das Verhältnis der Länge des pharyngalen Raums zum oralen Raum ändert.

Andererseits ist bekannt, dass vertikale Kehlkopflage und Grundfrequenz positiv miteinander korrelieren. So bietet Ohala (1973) in seinem Überblick neben eigenen Befunden (Ohala, 1972) weitere 38 (!) Quellen (wobei die früheste von 1773 stammt, wiederveröffentlicht in (Herries, 1974)) auf, die eine positive Korrelation der vertikalen Kehlkopfposition und der Grundfrequenz belegen sollen - und stellt nur eine einzige Quelle (P. Lieberman, 1970) dagegen, die für diesen Zusammenhang keine Evidenz aufzuweisen vermag. Neuere Evidenz für die positive Grundfrequenz-Larynxhöhe-Korrelation ist K. Honda, Hirai, Masaki und Shimada (1999), wo gezeigt wird, dass offenbar zumindest zu einer Absenkung der Grundfrequenz der Kehlkopf sinkt.

Wenn nun beide Geschlechter die gleiche Grundfrequenz produzieren, phonieren Männer eher in hoher, Frauen (und Kinder) eher in tiefer Lage. So kann man auch annehmen, dass die Versuchspersonen Maurers, die alle keine professionellen Sänger, die erlernt haben, die Kehlkopfposition relativ unverändert zu lassen, waren, mit unterschiedlich angehobenen Kehlköpfen produzierten, nämlich Männer mit einem angehobenen Kehlkopf, Frauen (und Kinder) mit einem tiefen Kehlkopf. Auf diese Weise ließe sich erklären, warum die gröbsten Unterschiede zwischen den tieferen Formanten verschwinden, denn bei Männern wird die natürlich vorgegebene größere Dimension des pharyngalen Raums verringert, bei Frauen jedoch nicht, so dass sie einander wieder ähnlich werden. So werden zumindest ansatzweise die Vokaltraktgrößen zwischen den Geschlechtern aneinander angeglichen, und es besteht weniger Bedarf, andere Quellen für die Variation niederer Formanten suchen zu müssen. Diese Erklärung lässt aber grundsätzlich die Befunde Maurers und seiner Kollegen weniger unplausibel erscheinen, als dies zunächst der Fall gewesen sein mag.

Jedenfalls muss man bei einer Untersuchung des Grundfrequenzeinsatzes damit rechnen, dass damit auch die vertikale Kehlkopflage variieren könnte, was andererseits wiederum Einfluss auf das Ansatzrohr und damit auf die Formanten haben kann. Dies wäre wieder einer der oben genannten „Nebeneffekte“.

Für Variationen des ersten Formanten, wenn diese denn aktiv gesteuert werden, müssen wir damit rechnen, dass – wie beim oben genannten *F1*-tuning der Sopranistinnen (Bresch & Narayanan, 2010) – der Öffnungsgrad des Kiefers variiert werden könnte. Dies wäre dann kein Kopplungseffekt einer anderen Artikulationsgeste, sondern wäre eine unabhängige artikulatorische Geste.

1.10 Ziele dieser Arbeit

In dieser Arbeit sollen Beiträge zur Diskussion um den Einfluss der Grundfrequenz auf die Vokalhöhe gegeben werden. Diese Beitrage teilen sich im Wesentlichen in drei Hauptteile:

- In einem ersten experimentellen Kapitel (2) wird die altersbedingte Variation der Grundfrequenz und der ersten drei Formanten in erwachsenen Sprechern des Englischen untersucht. Es ist klar, dass die altersbedingten Änderungen beim Übergang vom Kind zum Erwachsenen auf ein mehr oder weniger gleichzeitiges Wachstum sowohl des Kehlkopfes als auch des Ansatzrohres zurückzuführen sind, die in einer Absenkung sowohl der Grundfrequenz als auch aller Formanten resultieren. Es ist auch klar, dass altersbedingte physiologische Änderungen in erwachsenen Sprechern auftreten, und dass es Auswirkungen u.a. auf die Grundfrequenz und auf die Formanten gibt, aber es ist weniger klar, ob diese Änderungen ähnlich zu deuten sind wie die physiologischen Änderungen und deren akustische Konsequenzen beim Wandel vom Kind zum Erwachsenen (siehe hierzu den Literaturüberblick in Kapitel 2.1). Es ist vorstellbar, dass sich nur einzelne Parameter ändern, die als Konsequenz eine Art von natürlicher Perturbation der Produktion und/oder derer akustischen Konsequenzen erzeugen, für die kompensiert werden muss. Diese Fragestellung ist motiviert durch die Unklarheit darüber, welche Anteile der altersbedingten Variation durch soziophonetische, physiologisch bedingte, und möglicherweise kompensatorische Variation zu erklären und wie diese Anteile voneinander zu unterscheiden sind. Diese Fragestellung im Rahmen dieser Arbeit über den Einfluss der Grundfrequenz auf die Vokalhöhe zu untersuchen, liegt deshalb nahe, weil – zumeist unabhängig voneinander – zwei Effekte besonders häufig in der Literatur zu alternden Erwachsenenstimmen beschrieben werden, nämlich eine Variation des ersten Formanten und eine Variation der Grundfrequenz.

  Die Fragestellung wird anhand einer der ersten longitudinalen, mehrere Sprecher über mehrere Jahrzehnte verfolgenden akustischen Sprachdatenuntersuchung beantwortet werden.

- In einem zweiten experimentellen Kapitel (3) wird die Frage untersucht, welche Mittel Sprecher (des Deutschen) einsetzen, wenn ihr Sprachsignal akustisch perturbiert wird. Die Fragestellung wird anhand zweier unterschiedlicher Arten von Perturbation untersucht:

  – In einem ersten akustischen Perturbationsexperiment wird direkt das Vokalhöhenperzept der Sprecher beeinflusst, indem ihnen in Quasi-Echtzeit ihr eigenes Sprachsignal mit einem in der Frequenz verschobenenen erste Formanten präsentiert wird. Wir untersuchen - abweichend von den meisten anderen Studien dieser Art - nicht allein die Produktion des perturbierten Parameters, also in diesem Fall des ersten Formanten, sondern auch, ob sich die Produktion der Grundfrequenz ändert. Wir erwarten auf jeden Fall auch Änderungen im Parameter *f0*, da mit dem Automatismus der Intrinsischen Grundfrequenz zu rechnen

ist: wenn beispielsweise ein Sprecher, um für eine *F1*-Perturbation hin zu höheren Werten zu kompensieren, die Zungenhöhe anhebt, um tiefere *F1*-Werte zu produzieren, wird die höhere Zungenhöhe vermutlich eine durch biomechanische Kopplung beeinflusste höhere Grundfrequenz zur Folge haben - die Richtungen der produzierten Verschiebungen der beiden Parameter werden also immer gegenläufig sein.

Um einen von uns vermuteten *aktiven* Einsatz der Grundfrequenz zur Vokalhöhenfeinjustierung zu elizitieren, verwenden wir verschiedene Perturbationsstärken. Es ist bekannt, dass Kompensationen für akustische Perturbationen – im perturbierten Parameter gemessen – nicht nur praktisch nie vollständig sind, sondern oft ab einer bestimmten Perturbationsstärke geblockt sind (siehe den Literaturüberblick in Kapitel 2.2.1). Wir wollen untersuchen, ob in einem solchen Fall dann statt des ersten Formanten die Grundfrequenz verstärkt verändert wird.

– In einem zweiten Experiment untersuchen wir die kompensatorische Reaktion von Sprechern auf eine Grundfrequenzperturbation während einer sprachlichen Äußerung. Da bei einer aktiven Grundfrequenzänderung mit einer Variation der Kehlkopfhöhe zu rechnen sein wird, messen wir in diesem Fall zusätzlich zur produzierten Grundfrequenz auch die Formanten 1 bis 3.

Es stellt sich hier hauptsächlich die Frage, ob die Perturbation der Grundfrequenz, für die erwartet werden kann, dass für sie nicht vollständig kompensiert werden wird, stark genug ist, um eine Änderung des Vokalhöhenperzeptes hervorzurufen; sollte dies der Fall sein, ist mit einer kompensatorischen Produktionsverschiebung des ersten Formanten (und zwar in Gegenrichtung zu den erwarteten Effekten in allen Formanten, die durch eine mit der Grundfrequenzproduktion positiv korrelierten Kehlkopfhöhenvariation hervorgerufen werden könnten) zu rechnen.

- In einem dritten experimentellen Kapitel (4) wird ermittelt, inwieweit eine Grundfrequenzvariation in einem spektral ambigen Stimulus, der aus natürlicher Sprache resynthetisiert ist, die Vokalhöhenperzeption (insbesondere die Kategorisierung in vordere geschlossene bzw. halb-geschlossene Vokale) beeinflusst; der zweisilbige Stimulus wird hierzu in einen Trägersatz eingebettet. Variiert werden

  ... das Gespanntheitsmerkmal des Vokals. Es werden zwei Kontinua benutzt: *beten-bieten* vs. *Betten-bitten*. Wegen einiger Befunde, die darauf hinweisen, dass im Deutschen die Intrinsische Grundfrequenz in ungespannten Vokalen aktiver von Sprechern beeinflusst wird als in gespannten, wird hypothetisiert, dass Hörer im ungespannten Kontinuum sensitiver auf Grundfrequenzvariation reagieren

  ... der Typ der Grundfrequenzverschiebung: in einem Fall wird der Grundfrequenzverlauf der gesamten Äußerung *global* verschoben. Dies sollte eine Vokalhöhenbeeinflussung in allen Vokalen der Äußerung auslösen und somit eine Nutzung

der Grundfrequenz als auditiver Cue zur Erleichterung der Vokalhöhenzuordnung einschränken. Im zweiten Fall bleibt die Grundfrequenz auf die Silbe beschränkt, die den spektral ambigen Vokal enthält. Es ist damit zu rechnen, dass ein Teil dieser *lokalen* Grundfrequenzvariation von den Hörern einer intonatorischen Variation zugerechnet wird. Da der Stimulus jedoch spektral ambig ist, rechnen wir mit einer relativ deutlichen Nutzung der *lokalen* Grundfrequenzvariation als vokalintrinsischem Cue.

Generell soll also in dieser Arbeit getestet werden, ob Kovariation von Grundfrequenz und Formanten, und hierbei speziell des ersten Formanten, auftritt, und, falls eine Kovariation von *f0* und *F1* präsent sein sollte, ob es plausibel ist, dass diese wegen der Nutzung der Grundfrequenz als auditiver Cue zur Vokalhöhenperzeption zustande kommt.

Kapitel 2

Longitudinale Studien altersbedingter Veränderungen einiger ausgesuchter akustischer Korrelate von Quelle und Filter

2.1 Einführung

In diesem ersten experimentellen Kapitel wollen wir uns mit dem Altern der Stimmen Erwachsener und dessen Auswirkungen auf die Formanten, aber auch auf die Grundfrequenz beschäftigen. Wir wollen dies *auch* deshalb tun, da in vielen soziophonetischen Studien Kohorten mit Sprechern unterschiedlichen Alters benutzt werden, wobei die Zugehörigkeit zu einer Altergruppe gleichgesetzt wird mit der Zugehörigkeit zu einer Gruppe, die bestimmte (sozio-)phonetische Merkmale aufweist, also z. B. jene einer Varietät *vor* und *nach* einem stattgefundenen Lautwandel, wie z. B. in Hawkins und Midgley (2005), oder Variation im Gebrauch von dialektalen Merkmalen, wie z. B. in Müller, Harrington, Kleber und Reubold (2011). Dort, wo Merkmale untersucht werden, die eventuell auch von physiologisch bedingten Änderungen betroffen sind, ist dies natürlich nicht unproblematisch. So tun sich Watson und Munson (2007) denn auch schwer mit der Unterscheidung, wenn sie die „difference in vowel acoustics between older and younger adults, possibly related to age-related changes in vocal tract morphology“ (Seite 561) untersuchen, und zwar in zwei Alterskohorten, die einen Abstand von 50 Jahren aufweisen (Mitte 20 und Mitte 70) und dabei feststellen: „older adults retained the historically less-advanced more-back pronunciations of back-round vowels, as well as the less-extreme productions of /ɑ/ and /æ/, than the younger adults“ (Seite 563). Nun ist es zwar so, dass man diese Bewegungen (Frontierung hinterer Vokale und Öffnungsgraderweiterung für tiefe Vokale) als Lautwandel festgestellt hat (vgl. hierzu (Hillenbrand, Getty, Clark & Wheeler, 1995)), und doch ist nicht ganz klar, insbesondere bei den tieferen *F1*-Werten der älteren Sprecher, was davon möglicherweise Lautwandel ist und was physiologisch bedingt, denn eine *F1*-Absenkung,

die altersbedingt ist, wurde schon oft als konsistenteste und ausgeprägteste physiologisch bedingte Alterserscheinung beschrieben (Linville & Fisher, 1985a; Scukanec, Petrosino & Squibb, 1991; Xue & Hao, 2003), s. u. . Wie die Abbildung (2.1) aus Watson und Munson (2007, Seite 563) zeigt, trifft nämlich für praktisch alle Vokale zu, dass sie etwas „höher“ im Vokalraum liegen, also tiefere *F1*-Werte aufweisen. Zwar sind die tiefen Vokale offenbar mehr betroffen, dennoch erscheint es schwierig, anhand der *F1*-Werte quantifizierend eine Unterscheidung zwischen phonetischem und physiologisch bedingtem Wandel zu finden.[1]

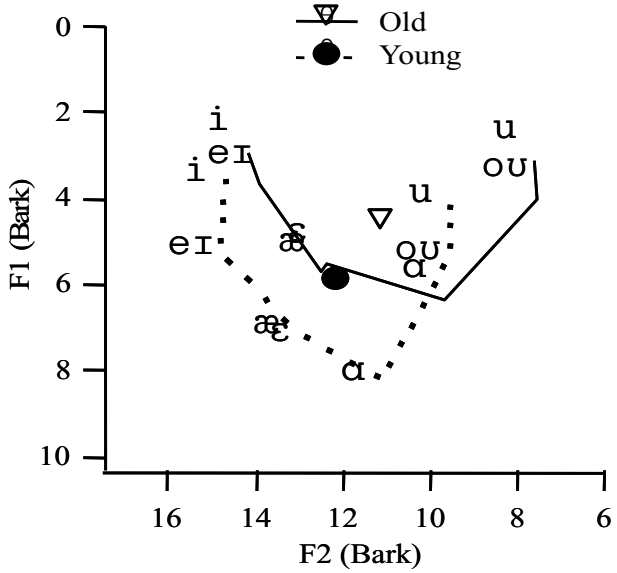

Abbildung 2.1: *Abbildung aus Watson und Munson (2007, Figure 1, Seite 563). Die originale Abbildungsunterschrift lautete: „Vowel space, expressed in F1/F2 bark values, for the young adults (dashed line) and for the old adults (solid line). The centroid value for the young adults is represented by the solid circle and the unfilled inverted triangle for the older adults.“*

Eckert (1997) gibt über die Problematiken der Untersuchungen soziolinguistischer Variablen mit verschiedenen Alterskohorten unter dem Titel „Age as a Sociolinguistic Variable“ einen recht umfassenden Überblick. Dort (Eckert, 1997, Seite 165) ist auch zu lesen:

> In general, though, adulthood has emerged as a vast wasteland in the study of variation. In sharp contrast to the year-by-year studies of children and adolescents, adults have been treated as a more or less homogeneous age mass.

Nun bezieht sich (Eckert, 1997) hierbei hauptsächlich auf soziolinguistisch relevante Variation. In der Erforschung physiologisch bedingter Variation jedoch gibt es durchaus viele Studien, die Alterskohorten von Erwachsenen miteinander vergleichen. Diese Studien zu altersbedingten physiologischen Änderungen im Erwachsenenalter sind also oft

[1] In der Tat tun das Watson und Munson (2007) auch nicht, sondern verwenden Maße der Euklidischen Distanz nach Bradlow (2002).

Querschnittsstudien, während Longitudinalstudien über die selben Sprecher zu diesen Fragestellungen, wie Verdonck-De Leeuw und Mahieu (2004) es feststellen, äußerst selten angetroffen werden können. Nun ist es aber durchaus wünschenswert, bei soziolinguistischen Fragestellungen auch die „Erwachsenen“ in unterschiedliche Gruppen einzuteilen (wie z. B. in Labov, Yaeger und Steiner (1972)), oder sogar Variation innerhalb eines Sprechers zu messen, so wie es Harrington und Kollegen (Harrington, 2006, 2007; Harrington, Palethorpe & Watson, 2000a, 2007b) in mehreren Studien zur Teilnahme der britischen Königin Elisabeth II am Lautwandel, der in der Standardaussprache der Standard Southern British-Varietät, der Received Pronunciation, stattgefunden haben soll, getan haben. Andererseits ist es gleichfalls wünschenswert, von den möglicherweise zu sehr durch Sprecherspezifika beeinflussten Querschnittsstudien zu physiologischen Alterseffekten wegzukommen und auch hier eine longitudinale Beobachtungsweise zu den Änderungen innnerhalb eines Sprechers anzustreben. Auf diese Weise ließe sich der Einfluss des einen Faktors (z. B. physiologischer Änderungen im Alter) eher von denen des anderen (z. B. Lautwandel) abgrenzen – wenn auch für soziophonetische Daten relative Maße wie – im Bereich der Formanten von Vokoiden – Euklidische Distanzen einzelner Vokale zu Ankerpunkten, die aus der Verteilung der Tokens zweier anderer Vokaltypen desselben Sprechers bestehen und so die relative Lage der Vokaltokens innerhalb des Vokalraums des Sprechers bestimmen lassen, mit einiger Sicherheit ohnehin die bessere Wahl sind (siehe hierzu beispielsweise Harrington (2006); Harrington et al. (2008), wo diachronisch stabile Vokale als Ankerpunkte für Vokale, bei denen eine Vokalverschiebung durch Lautwandel vermutet wurde, benutzt wurden, um die relative Position zwischen diesen Ankervokalen zu quantifizieren).

Alter und dessen Einfluss auf Grundfrequenz und Formanten Im Folgenden sollen also die Befunde aus der Literatur zu physiologisch bedingten Änderungen, die mit Alter zusammenhängen, beschrieben werden. Einen umfassenden Überblick zur Literaturlage zum *vocal aging* findet man im gleichnamigen Buch von Sue Ellen Linville (Linville, 2001), wobei natürlich festzuhalten ist, dass zum Themenbereich – im Zeitalter zunehmend steigender Lebenserwartung und der damit verbundenen Popularität des Themas – zahlreiche Studien nach dem Erscheinungsdatum dieses Standardwerkes hinzugekommen sind. Leider gilt immer noch, was Hollien (1987) schon vor einem Vierteljahrhundert, als er „Old voices: what do we really know about them?“ (so der Titel seiner Schrift) fragte, so ausgedrückt hat: „a rather substantial (but, perhaps, somewhat *unorganized*) literature on the subject is becoming available“ (Seite 12, Hervorhebung von mir).

Die Auswahl der hier zu beschreibenden Literatur soll sich auf Daten begrenzen[2], die relevant sind für die Vokalperzeption, also Formanten und auch Grundfrequenz. Relativ wenige Studien über die Altersstimme beschäftigen sich mit Formanten. Einige Studien der – wie gesagt – hauptsächlich als Querschnittsstudien designten Untersuchungen finden einen Alterseffekt auf Formanten, wobei der konsistenteste dieser Effekte eine Absenkung des ersten Formanten ist (Linville & Fisher, 1985a; Linville, 1987a; Linville & Rens, 2001;

[2]Das heißt zum Beispiel, dass auf die recht zahlreiche Literatur zu Perturbationsmaßen der Grundfrequenz – also Quantifizierungen von Effekten wie *jitter* und *shimmer* – nicht eingegangen werden wird.

Scukanec et al., 1991; Xue & Hao, 2003). Einige Studien kommen zu dem Befund, es mit einer generellen Verkleinerung der Vokalraums zu tun zu haben (Rastatter, McGuire, Kalinowski & Stuart, 1997). Dieser Zentralisierungseffekt ist bei Männern offenbar stärker ausgeprägt als bei Frauen. Linville nimmt diesen in ihr „blended model of vocal tract resonance changes with aging“ auf (siehe Abbildung 2.2 und Linville (2001, Kapitel 10, insbesondere Seiten 179 bis 183)), das zwischen Frauen und Männer unterscheidet; gemeinsam ist beiden Geschlechtern in diesem Modell ein Absinken der Formantfrequenzen, während sie die Zentralisierungstendenz nur den Männern zuschreibt. Auch Rastatter und Jacques (1990) fand vokal- und geschlechtsabhängige Unterschiede zwischen den altersbedingten Änderungen bei den Formanten 1 und 2; wegen der relativ geringen Sprecheranzahl (10 Sprecher pro Gruppe) und nicht vorgenommenen Normalisierung ist dieser Befund allerdings schwierig zu interpretieren. Doch auch in dieser Studie ist der Einfluss des Alters bei männlichen Sprechern am stärksten bei *F1* ausgepägt - bei Frauen hingegen für *F1* und *F2* gleichermaßen. In Linville und Rens (2001) wird ebenfalls berichtet, dass – neben einem bei beiden Geschlechtern ausgeprägten altersbedingten Absinken des ersten Gipfels eines *long-term-average-spectrums* (LTAS) – der zweite und dritte Gipfel des LTAS bei Frauen signifikant, bei Männern tendenziell sinken.

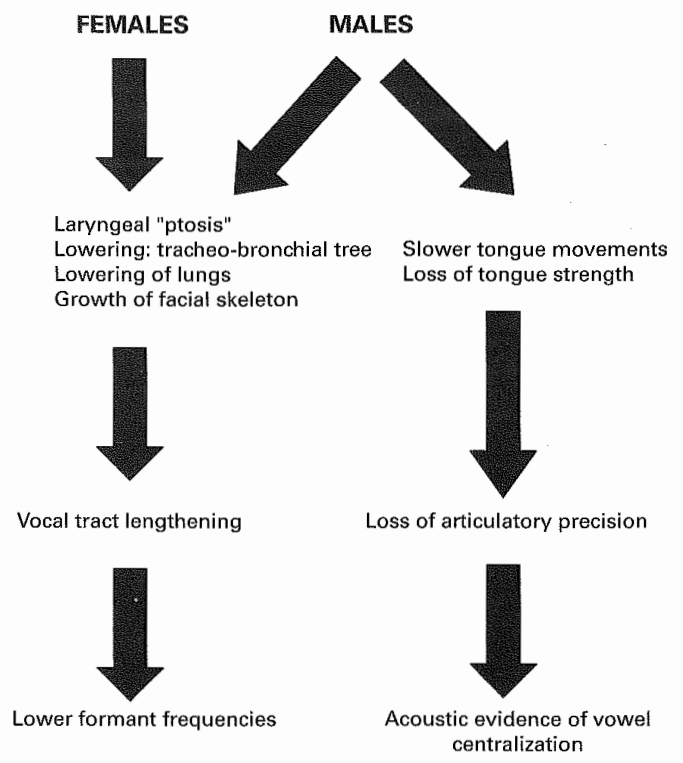

Figure 10–3. Flowchart illustrating the blended model of vocal tract resonance changes with aging.

Abbildung 2.2: *Linvilles „blended model of vocal tract resonance changes with aging“, zitiert nach Linville (2001, Seite 182).*

Dieser Abfall des ersten und eventuell weiterer Formanten wurde oft den bereits zuvor beschriebenen physiologischen Änderungen des Vokaltraktes angerechnet, da vermutet wurde, dass diese Änderungen zu einer Verlängerung des Vokaltraktes führen sollten, wie schon Ferreri (1959) (zitiert nach Linville (2001, 2002); Rastatter und Jacques (1990); Hoit und Hixon (1992); Melcon, Hoit und Hixon (1989) und vielen weiteren) anhand anatomischer Messungen festzustellen glaubte. In Corso (1975) und T. Cohen und Gitman (1959) (letztere zitiert nach Rastatter und Jacques (1990)) werden Muskelatrophie und Dehnung der Kehlkopfmuskulatur beschrieben, in Israel (1968, 1973) das auch im Erwachsenenalter fortgesetzte Wachstum der Strukturen des Gesichtsschädels. Einen weiteren Überblick hierzu bietet wieder Linville (2001, Kapitel 3 über laryngale Änderungen, Kapitel 4 über Änderungen der supraglottalen Strukturen). Ob all diese Effekte tatsächlich zu einer Absenkung des Kehlkopfs (relativ zu den Wirbeln der Halswirbelsäule) führen, ist allerdings fraglich. So zitiert das Standardwerk von Zemlin (1998) die Arbeit von Wind (1970), der eine vertikale Absenkung des Kehlkopfs im Laufe des Erwachsenenalters berichtet. Gemessen an der Untergrenze des Cricoids fällt der Larynx vom sechsten Halswirbel bei 20jährigen bis zur Höhe des siebten Halswirbels bei 80jährigen ab.[3] Flügel und Rohen (1991) maßen anhand von sagittalen Röntgenbildern von 116 Personen im Alter zwischen 12 Tagen und 71 Jahren. Für das Erwachsenenalter konnten dort keine konsistenten altersbedingten Lageveränderungen des Larynx' festgestellt werden; die tiefste relative Lage wurde für einen 32jährigen gefunden.

Dies schließt natürlich weiteres Wachstum des Gesichtsschädels und eine daraus entstehende Verlängerung des Ansatzrohrs nicht aus. Xue und Hao (2003) fanden in ihrer volumetrischen Messung mittels *acoustic reflection technique* (ART), einer Technik, die erlaubt, auf zuverlässige Weise (Fredberg, Wohl, Glass & Dorkin, 1980; Marshall et al., 1993) die räumlichen Abmessungen von Hohlräumen zu ermitteln, keine altersbedingten Unterschiede der Länge des gesamten Vokaltraktes beim Vergleich 38 junger (18-30 Jahre alter) und 38 alter (62-79jährige) Versuchspersonen beiderlei Geschlechts (auch nicht in der Vorstudie mit 22 Frauen in Xue, Jiang, Lin, Glassenberg und Mueller (1999)), dafür aber einen Anstieg des Volumens des gesamten Vokaltraktes, und eine Vergrößerung des Volumens und eine Längung des oralen Raums bei den älteren Versuchspersonen.[4]

Etwas ausführlichere Daten als für Formanten liegen für die altersbedingten Veränderungen der Grundfrequenz vor. Bei Frauen ist es offenbar der Fall (mit wenigen Ausnahmen wie den Daten in Biever und Bless (1989)), dass die Grundfrequenz sinkt (Baken, 2005; Linville, 1996; Nishio & Niimi, 2008; Pegoraro Krook, 1988; da Silva, Master, Andreoni, Pontes & Ramos, 2010; Torre III & Barlow, 2009; Stoicheff, 1981). Dieser Befund aus Querschnittsstudien oder Kompilationen verschiedener Studien wird auch durch longitudinale Studien innerhalb derselben Personen bestätigt (Harrington et al., 2007a; Mwangi et al., 2009). Dies gilt auch für De Pinto und Hollien (1982), die nach etwas über fünfunddreißig

[3] Bedauerlicherweise war es nicht möglich, an die Quelle, also an Winds Dissertation, zu gelangen, so dass nicht bekannt ist, wieviele Personen ausgemessen wurden, und mit welcher Methodik dies geschah.

[4] Zur Erinnerung: In der gleichen Studie (Xue & Hao, 2003) wurden – wie oben erwähnt – bei beiden Geschlechtern altersbedingte Formantänderungen hauptsächlich bei F1 festgestellt, mit einem tieferen ersten Formanten im höheren Alter.

Jahren die gleichen 11 Frauen (aus einer 1945 28 Frauen starken Gruppe 18- bis 19jähriger Frauen) denselben Text nochmals lesen ließen und einen signifikanten Abfall der Grundfrequenz feststellten. Weitere 13 Jahre später ließen Russell, Penny und Pemberton (1995) nochmal einige Frauen aus der Ursprungsgruppe den gleichen Text lesen und stellten das gleiche Resultat fest. Nur sechs Frauen in dieser Studie waren auch Teilnehmerinnen der Studie in De Pinto und Hollien (1982). Diese wurden gesondert mit den Messungen aus De Pinto und Hollien (1982) verglichen, aber es wurden keine weiteren Unterschiede der Grundfrequenz festgestellt. Bei sehr alten Frauen soll es jenseits des Alters von 90 Jahren wieder zu einem Anstieg der Grundfrequenz kommen können (Max & Mueller, 1996), wohingegen andere Studien keinen Anstieg feststellen konnten (Awan & Mueller, 1992).

Bei Männern scheint die Datenlage zunächst etwas uneinheitlicher zu sein, da manchmal berichtet wird, *f0* ändere sich nicht signifikant (Ramig & Ringel, 1983; Verdonck-De Leeuw & Mahieu, 2004) (bei der letzten Quelle hier allerdings nur in einem 5-Jahres-Abstand gemessen und abhängig von Gewohnheiten wie Rauchen), sinke (Benjamin, 1981; Decoster & Debruyne, 2000; Guimarães & Abberton, 2005; Harrington et al., 2007a) oder steige an (Chen, 2007; Harnsberger, Wright & Pisoni, 2008; Mysak & Hanley, 1958; Torre III & Barlow, 2009). Solche sich scheinbar widersprechenden Ergebnisse könnten daher rühren, dass die Grundfrequenz bei Männern zunächst sinkt und dann wieder ansteigt, wie Kompilationen von Studien (Baken, 2005; Brown et al., 1991; Linville, 1996, 2001) oder Studien, die tatsächlich jedes Lebensjahrzehnt zwischen 20 und 90 abdecken (wie Hollien und Shipp (1972)), vermuten lassen.

Abbildung 2.3 zeigt kompilierte Daten, zusammengefasst in Baken (2005); Brown et al. (1991); Linville (1996, 2001) für Frauen und Männer, sowie von Hollien und Shipp (1972) selbst erhobene Daten von Männern. Gemein ist diesen Abbildungen, dass sie für Männer zunächst ein Absinken der Grundfrequenz, gefolgt von einem Wiederanstieg der Werte, zeigen. Leichte Unterschiede bestehen darin, dass bei Linville und Baken die Grundfrequenz schon ab einem relativ frühem Alter (bei beiden in etwa um das 40ste Lebenjahr herum) wieder deutlich ansteigt und somit mehr oder weniger U-förmige Verläufe zu sehen sind, während bei Brown und Kollegen und bei Hollien und Shipp ein eher W-förmiger Verlauf vorherrscht, also die Grundfrequenz ein (lokales) Minimum bei 40 bis 45 Jahren erreicht, danach ansteigt bis etwa Mitte fünfzig, um danach wieder abzusinken bis etwas Mitte sechzig, um dann endgültig wieder anzusteigen zu Werten, wie sie bislang im Erwachsenenalter von Männern nicht zu finden waren. Bei Baken ist zu beachten, dass er auch männliche Kinderstimmen mit einschließt und erst dadurch in seinen Daten der zunächst stattfindenden Abfall der Werte deutlich wird; es ist nicht klar, ob seine Daten, nur für die Sprecher zwischen dem 20sten und dem 90sten Lebensjahr modelliert, überhaupt signifikante Änderungen aufweisen würden. Jedenfalls bestätigen die hier gezeigten Daten im allgemeinen eine zunächst fallende Grundfrequenz und danach einen Wiederanstieg in höheren Lebensaltern jenseits der 40.

Was die weiblichen Grundfrequenzen in diesen Überblicken angeht, so ist festzustellen, dass Brown et al. (1991) eher keine Änderung im Erwachsenenalter darstellen, Baken einen eher sanften, kontinuierlichen Abfall und eventuell einen leichten Anstieg im Alter über 80,

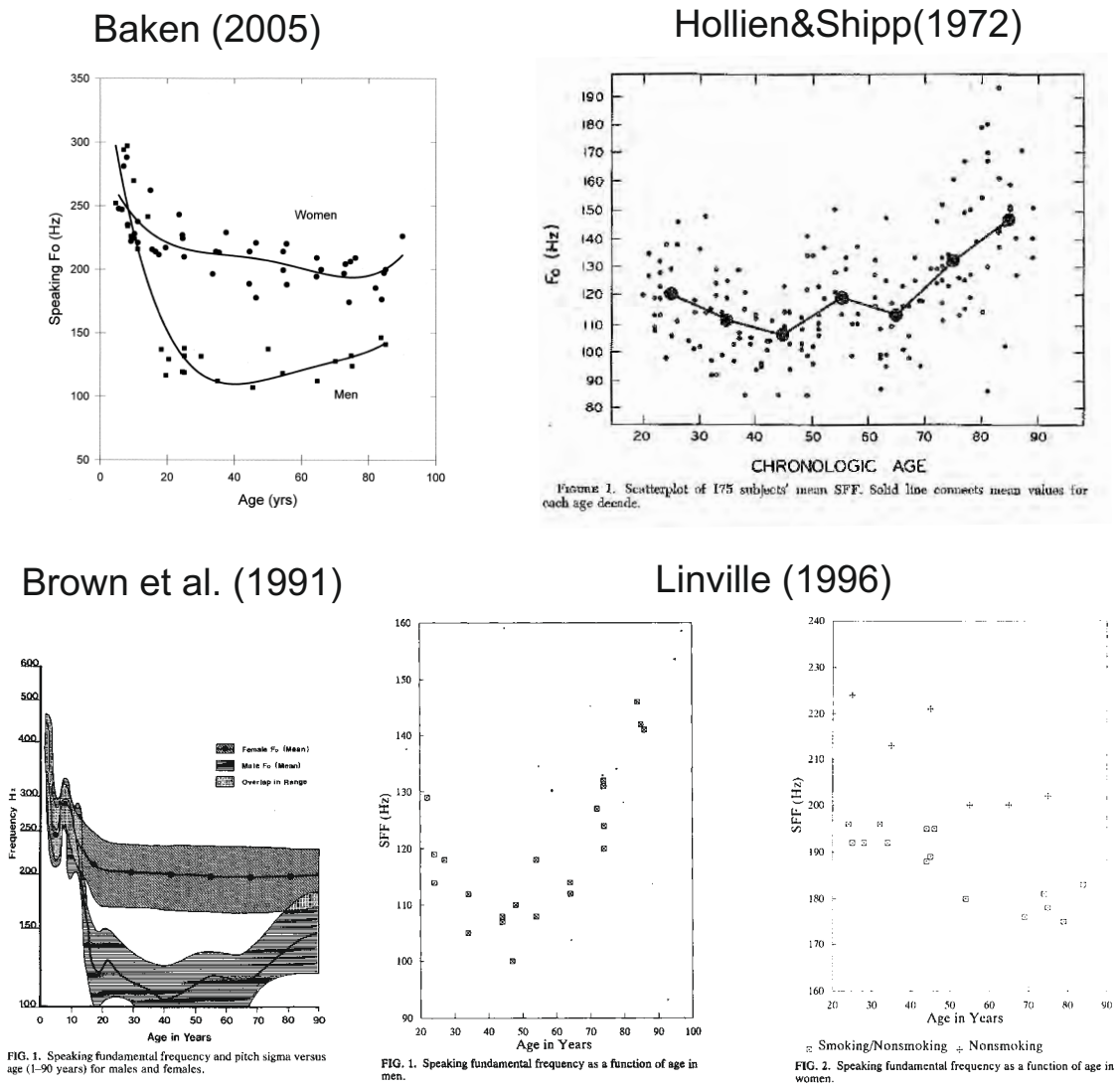

FIG. 1. Speaking fundamental frequency and pitch sigma versus age (1–90 years) for males and females.

FIG. 1. Speaking fundamental frequency as a function of age in men.

FIG. 2. Speaking fundamental frequency as a function of age in women.

Abbildung 2.3: *Mittlere Grundfrequenzen über Kohorten von Sprechern zu verschiedenen Altern, aus Baken (2005), Brown et al. (1991), und Linville (1996) (so auch in Linville (2001) (dort links Männer, rechts Frauen, diese unterteilt in Raucherinnen und Nichtraucherinnen)), sowie aus (Hollien & Shipp, 1972). Man beachte die unterschiedliche Auswahl der Altersabschnitte und die unterschiedlichen Skalierungen. Nur (Hollien & Shipp, 1972) ist eine eigenständige Untersuchung mit 175 Männern im Alter zwischen 20 und 89, die kontrolliertes Material lasen (die rainbow passage).*

und Linville einen eher stufenweisen Abstieg, mit einer „Stufe“ zwischen 50 und 60.[5]

Diesen eher abrupten Abfall der Werte führt Linville deshalb auch auf hormonelle Än-

[5]Man beachte bei Linvilles Daten, dass sie zwischen Raucherinnen und Nichtraucherinnen in ihrer Abbildung unterscheidet. Für Raucherinnen sind tiefere Grundfrequenzen auszumachen; für beide Gruppen ist aber der stufenartige, abrupte Abfall der Werte zwischen 50 und 60 zu beobachten.

derungen zurück, die Einfluss auf die Stimmlippen und somit deren Schwingungsverhalten haben, wie sie in Abitbol, Abitbol und Abitbol (1999) beschrieben werden. Bei den angedeuteten hormonalen Änderungen handelt es sich um ein dramatisches Absinken des Spiegels der Östrogene während der Menopause, so dass Linvilles Kompilation damit gut erklärt werden kann. Auch die Daten in Baken (2005) zeigen einen etwas ausgeprägteren Abstieg der Werte bei diesem Alter, welcher auch mit der abrupten Spiegeländerung der Östrogene erklärt werden könnte . Bei Männern hingegen sinkt der Testosteronspiegel langsam, aber kontinuierlich mit dem Lebensalter, was, wie Gugatschka et al. (2010) zeigen, aber nicht in Beziehung gesetzt werden kann mit Grundfrequenzveränderungen - obschon für jüngere Männer im Falle einer Testosteron-Ersatz-Therapie deutliche Auswirkungen auf die Grundfreqeunz festgestellt werden konnten (King, Ashby & Nelson, 2001) und Evans, Neave, Wakelin und Hamilton (2008) sogar eine negative Korrelation zwischen dem über den Tag hinweg schwankenden Testosteronspiegel und der Grundfrequenz finden. Auch im männlichen Körper wirken aber Östrogene, und für diese konnten in der Alterstimmenuntersuching in Gugatschka et al. (2010) eindeutige Einflüsse auf die Grundfrequenz nachgewiesen werden - bei niedrigen Östrogenwerten ist die Grundfrequenz höher (!).

Es kann an dieser Stelle nicht entschieden werden, welches Hormon nun genau die altersbedingten Grundfrequenzänderungen auch bei Männern beeinflusst; fest steht nur, dass einzelne Sexualhormone bzw. die Verteilung der Östrogene und der Androgene, welche sich mit zunehmendem Alter zwischen Männern und Frauen angleicht, die Stimmlippen beeinflusst, indem der hormonelle Einfluss die Stimmlippen entweder (bei Frauen) anschwellen oder (bei Männern) dünner werden lässt. Wie bekannt, ist die Frequenz der Stimmlippenschwingung mit der Masse der Stimmlippen negativ korreliert, d.h. bei Frauen würde man also durch die beobachtete Massenzunahme (durch Ödeme) eine sinkende, bei den Männern durch Verringerung der Masse (durch Atrophie) der Stimmlippen eine steigende Grundfrequenz erwarten, was Honjo und Isshiki (1980) auch so finden. Hollien (1987) schlägt deshalb auch als Alterungsmodell das *male-female coalescence model* vor, also die Wiederangleichung von Männer- und Frauenstimmen im höheren Alter, so wie auch vor der Pubertät Jungen- und Mädchenstimmen relativ ähnlich gewesen seien. Hierbei bezieht sich Hollien (1987) aber hauptsächlich auf die Grundfrequenz; von einer Rücknahme der zweiten Kehlkopfsenkung bei männlichen Personen während der Pubertät (und der dadurch entstandenen tieferen Formantwerte) kann ja auch nicht ausgegangen werden.

Ein weiterer Befund, der möglicherweise zusätzlich erklärt, warum bei Männern im hohen Alter *f0* so deutlich ansteigt, ist in Hirano, Kurita und Sakaguchi (1989) zu finden: Die Stimmlippen alter Männer jenseits der siebzig sind kürzer als jene jüngerer Männer, während bei Frauen nur sehr geringe altersbedingte Unterschiede in der Länge der Stimmlippen zu finden sind (Hirano, Kurita & Sakaguchi, 1988), was wiederum gut zu den Daten in der Abbildung aus (Brown et al., 1991), also den nur gering ausgeprägten *f0*-Änderungen, passt.

Ein Zusammenhang könnte bestehen zwischen dem Wiederanstieg der Grundfrequenz bei Männern und sehr alten Frauen und dem in da Silva et al. (2010) beschriebenen Abfall des Schalldruckpegels bei alternden Sprechern; die Autoren beziehen sich auf Titze und Sundberg (1992) und deren Befund, dass bei gleichbleibendem subglottalen Druck die

Intensität positiv mit der Grundfrequenz korreliert; Erhöhung der Grundfrequenz könnte also das Erreichen der notwendigen Intensität erst ermöglichen, da der subglottale Druck im höheren Alter rapide abnimmt (da Silva et al., 2010); eine Erhöhung der Grundfrequenz könnte aktiv eingesetzt werden, um kompensatorisch dem Abfall des subglottalen Drucks entgegenzuwirken und einen ausreichend hohen Schalldruckpegel zu erreichen.

Unterschiede zwischen Generationen[6] Wie beschrieben, stammten die meisten der hier dargestellten Befunde aus Querschnittsstudien; dies könnte, wie bereits erwähnt, gewisse Überschneidungen physiologischer Alterserscheinungen mit soziolinguistischer Variabilität zur Folge haben, wovon allerdings auch longitudinale Studien nicht frei sind, siehe zu letzterem z. B. Harrington et al. (2000a); Harrington, Palethorpe und Watson (2000b); Harrington (2006); Harrington et al. (2007b); Harrington (2007, über Lautwandel innerhalb der britischen Königin) oder Sankoff und Blondeau (2007). Als Beispiel für generelle Generationenunterschiede sei Pemberton, McCormack und Russell (1998) genannt; dort wurden die gleichen Archivaufnahmen benutzt wie für den longitudinalen Alterstimmenvergleich in De Pinto und Hollien (1982) und Russell et al. (1995). Die Aufnahmen aus dem Jahre 1945 wurden aber nicht mit jenen der selben Sprecherinnen im höheren Alter verglichen, sondern mit den Stimmen von ebenfalls 28 jungen Frauen (Alter 18 bis 25), die im Entstehungsjahr der Studie (1993) das gleiche College besuchten und das gleiche Alter hatten wie die ursprünglichen Sprecherinnen im Jahr 1945. Es wurde eine signifikant tiefere mittlere Grundfrequenz festgestellt (ein Absinken von 229 Hz im Jahr 1945 auf 206 Hz im Jahr 1993, was 1.83 Halbtönen entspricht). Die Werte für die unterschiedlichen Generationen entsprechen in etwa denen aus dem Literaturüberblick in Pemberton et al. (1998) für junge Frauen in industrialisierten Gesellschaften. Da in ihrer Studie nicht nur das gelesene Material kontrolliert worden war, sondern auch grundfrequenzerniedrigende Faktoren wie Rauchen (alle 56 Versuchspersonen waren zum Zeitpunkt der Aufnahme Nichtraucherinnen und hatten vorher auch nicht geraucht) sowie die Einnahme von Medikation, die 1945 noch nicht verfügbar war (Asthmasprays bzw. Kontrazeptiva auf Hormonbasis (die sogenannte „Antibabypille“)), und diese Faktoren daher nicht Auslöser für den gefundenen Unterschied sein konnten, schlossen die Autoren auf eine Beeinflussbarkeit der Grundfrequenz durch soziale Faktoren, d.h. sozusagen durch eine kulturell bedingte „Mode“ für tiefere Frauenstimmen, die auch durch die Medien verbreitet werden würde.

Daneben gibt es aber auch andere Gründe, skeptisch zu sein, was Vergleiche zwischen Generationen betrifft, denn ist zum Beispiel vorstellbar, dass in den hier untersuchten Populationen sich grundsätzliche biologische Unterschiede zwischen den Generationen zei-

[6] Dem Leser mag dieser Abschnitt als ein etwas zu weit gehender Exkurs erscheinen. In der Tat wurde ein ähnlicher Absatz aus einem ersten Entwurf zu (Reubold, Harrington & Kleber, 2010) vor der Absendung des Typoskripts wieder gelöscht. Einer der Gutachter warf allerdings auch die Behauptung auf, er könne sich nicht vorstellen, dass die Größe des Kehlkopfs und des Ansatzrohrs der hier vorgestellten Personen nicht beeinflusst worden sein sollte durch die schlechteren Lebensbedingungen während ihrer Kindheit und Jugend, wie sie durch die Wirtschaftskrise der 20er Jahre und die Essensrationierung während des Zweiten Weltkrieges zu erwarten sein müssten. Für Reubold et al. (2010) wurde der Abschnitt dennoch nicht wieder aufgenommen; der Vollständigkeit halber ist er hier enthalten.

gen. Ein Effekt kann beispielsweise im Bereich der durchschnittlichen Körpergröße und des durchschnittlichen Körpergewichts angenommen werden – und man könnte eine gewisse Korrelation zwischen Körpergröße und „Tiefe“ der Stimme annehmen, wenn man die Daten aus Chambers, Cleveland, Kleiner und Tukey (1983) in Abbildung 2.4 über die Größenverteilung von 283 Sängern der New York Choral Society, aufgeteilt nach Gesangsstimmtyp, betrachtet, auch wenn man recht viel Überlappung feststellen muss und die Beziehung zwischen der Klassifikation der Gesangsstimme möglicherweise nicht so klar mit der Grundfrequenz beim Sprechen korreliert ist (Larsson & Hertegård, 2008), wie man annehmen möchte.

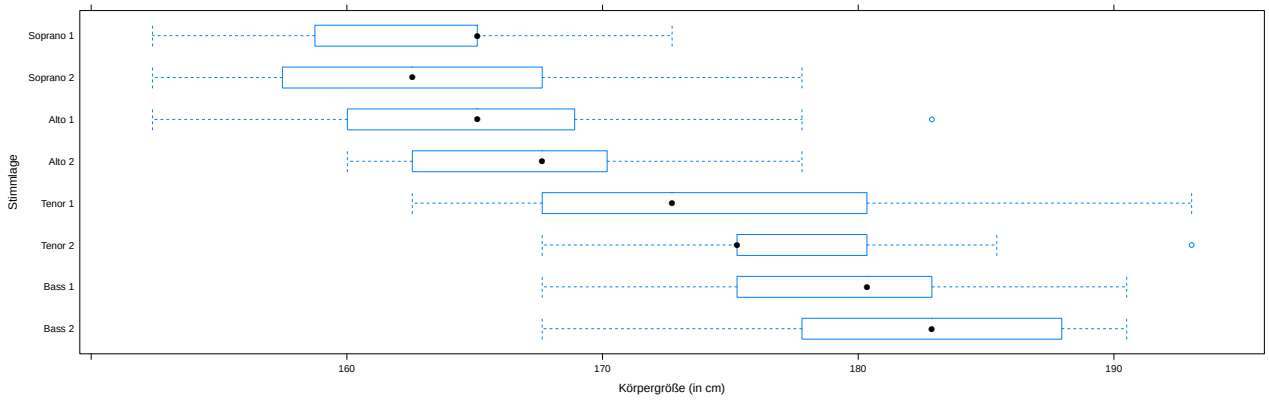

Abbildung 2.4: *Körpergrößeverteilung der Gesangsstimmtypen, mit abnehmender mittleren Gesangsstimmtonhöhe. Daten zitiert aus Chambers et al. (1983).*

Was die Zunahme der durchschnittlichen Körpergröße und des Gewichts angeht, zeigten Ogden, Fryar, Carroll und Flegal (2004) für Daten, die in den Vereinigten Staaten von Amerika zwischen 1960 und 2002 erhoben wurden, bei Erwachsenen eine Zunahme des durchschnittlichen Körpergewichts um 24 Pfund und eine Erhöhung der Körpergröße um 3 Zentimeter bei Männern und um 2 Zentimeter bei Frauen. Diese allgemeine Erhöhung der Werte betrifft offenbar alle Bevölkerungsgruppen, also beide Geschlechter, alle ethnischen Gruppen und alle Altersgruppen. Da das Gewicht dramatischer zugenommen hat als die Körpergröße, stieg auch der sogenannte *Body-Mass-Index*[7] von circa 25 auf circa 28. Für 10 westeuropäische Länder zeigten Cavelaars et al. (2000) für Geburtskohorten im 5-Jahres-Abstand eine generelle Zunahme der Durchschnittsgröße im Erwachsenenalter von 0,8 Zentimetern bei Männern und von 0,4 Zentimetern bei Frauen pro Kohorte im ausgehenden zwanzigsten Jahrhundert. Diese Unterschiede können mit den im Verlauf des zwanzigsten Jahrhunderts veränderten Lebensbedingungen in den industrialisierten Staaten, in denen die Daten erhoben worden waren, erklärt werden.

Es ist allerdings durchaus umstritten, ob dieses Wachstum und die Gewichtszunahme überhaupt Auswirkungen haben können oder nicht. Künzel (1989) und van Dommelen

[7] $Body - Mass - Index = \frac{Körpermasse\ [in Kilogramm]}{(Körpergröße\ [in Metern])^2}$

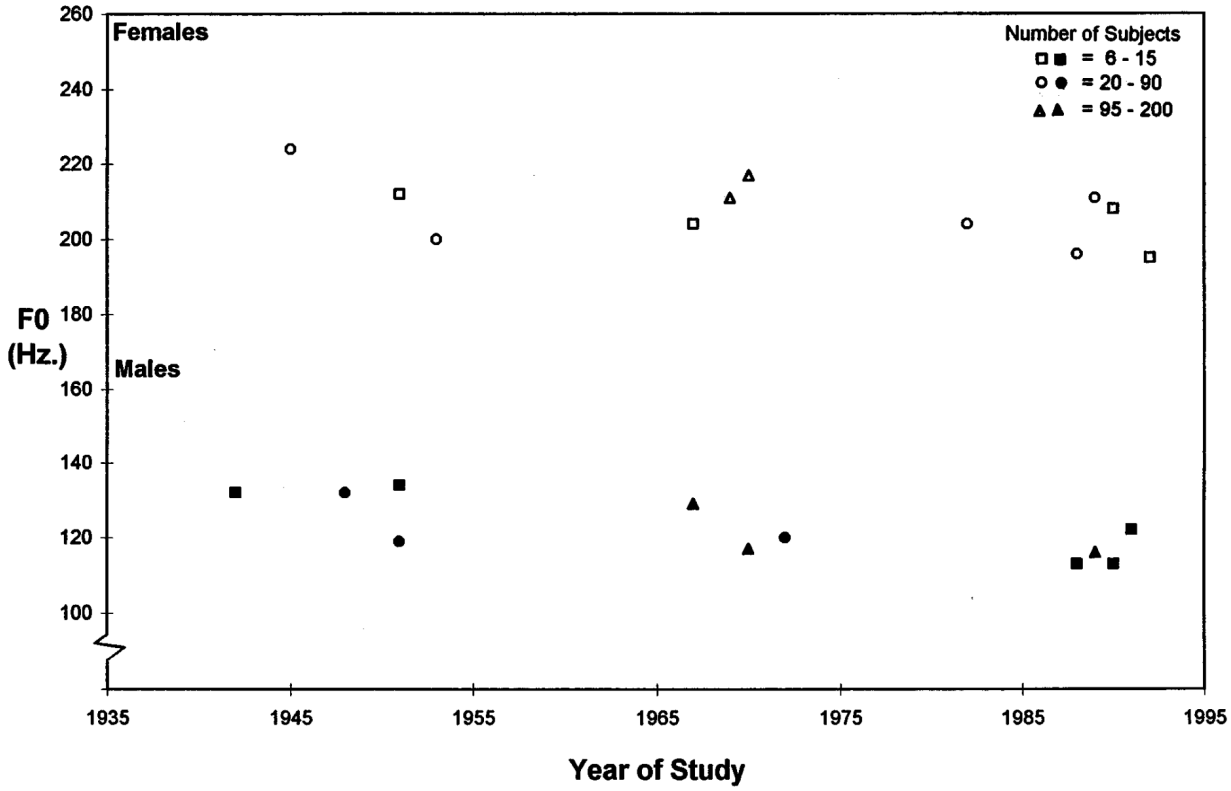

FIG. 2. Plot of mean speaking fundamental frequency data from 17 studies (American/European subjects only, drawn from de Pinto and Hollien, 1982; Fitch, 1990; Fitch and Holbrook, 1970; Hanley, 1951; Hanley and Snidecor, 1966; Higgins and Saxman, 1991; Hollien and Jackson, 1973; Hollien and Paul, 1969; Hollien and Shipp, 1972; Hollien *et al.*, 1982; Krook, 1988; Künzel, 1989; Linke, 1953/1973; Philhour, 1948; Pronovost, 1942; Snidecor, 1951; Yamazawa and Hollien, 1992). Note that group size is classified as "small" (squares), "medium" (circles), or "large" (triangles).

Abbildung 2.5: *Mittlere f0 im Verlauf des zwanzigsten Jahrhunderts, zitiert nach Hollien et al. (1997, Seite 2986), mit originaler Abbildungsunterschrift.*

(1993) zeigten eher, dass *f0 nicht* mit der Körpergröße bzw. dem Körpergewicht korreliert, während Graddol und Swann (1983) eine solche Korrelation zumindest bei Männern feststellt. Hollien et al. (1997) zeigen zwar ein Absinken der Grundfrequenz als allgemeinen Trend (siehe Abbildung 2.5), weisen aber auf methodische Unterschiede in den Studien, aus denen sie die abgebildeten Daten entnommen haben, hin (Anzahl der Raucher und Nichtraucher, unterschiedliche soziale Gruppen, kleine Unterschiede des Versuchspersonenalters, Gruppengrößen etc.) und folgern daraus, dass, sofern dieser Trend überhaupt existiere, er im letzten Viertel des zwanzigsten Jahrhunderts zum Erliegen gekommen sei. González (2004) zeigt eine nur gering ausgeprägte Beziehung zwischen Körpergröße und Formantlagen. Eine heftig geführte Diskussion entstand um die Ergebnisse von Lass und Kollegen (Lass & Davis, 1976; Lass, Beverly, Nicosia & La Simpson, 1978; Lass et al., 1979; Lass, Phillips & Bruchey, 1980), dass akustische Merkmale (wie die Grundfrequenz und Formantlagen) von Hörern genutzt werden könnten, um Gewicht und Größe des Sprechers zuverlässig abzuschätzen, was von J. R. Cohen, Crystal, House und Neuburg (1980) stark

angezweifelt wurde (siehe eine Gegenkritik hierzu wiederum auch in Lass (1981)). Gemein ist diesen Befunden, so unklar sie teilweise auch sein mögen, dass sie eher in Gegenrichtung zu den bisherigen Befunden zur Altersstimme verlaufen: das vermutete Absinken der Grundfrequenz bei älteren Frauen beispielsweise könnte also davon verdeckt werden, dass es in jüngeren Kohorten von Sprecherinnen „Mode“ geworden ist, mit tieferer Stimme zu sprechen, oder dass eine Zunahme von Körpergröße und -gewicht bei jüngeren Generationen zu einer Absenkung der Grundfrequenz und/oder der Formanten führt. Wie wir gesehen haben, ist es ja der Fall, dass es z. B. bezüglich der Grundfrequenz der Frauen umstritten ist, inwieweit bei ihnen altersbedingt die Grundfrequenz sinkt (z. B. in Linville (2001)) oder nicht (z. B. in Brown et al. (1991) oder in Biever und Bless (1989)).

Longitudinale Studien zum Alter Einige der genannten Befunde zur Altersstimme entstammten aber auch longitudinalen Studien, also aus Untersuchungen anhand der selben Sprecher zu unterschiedlichen Lebensaltern, so z. B. Decoster und Debruyne (2000); Harrington (2006); De Pinto und Hollien (1982); Russell et al. (1995), wobei also Veränderungen über Generationen hinweg ausgeschlossen sein sollten, vielleicht mit der Ausnahme der bereits erwähnten Befunde zur Teilnahme einzelner Sprecher an soziophonetischen Veränderungen. Zu erwähnen wäre an dieser Stelle im Bereich der longitudinalen Studien zu altersbedingten Veränderungen noch Endres, Bambach und Flosser (1971), welche fanden, dass die Mittelwerte über die ersten 4 Formanten bei vier Männern und zwei Frauen – im Abstand von circa eineinhalb Jahrzehnten gemessen – sanken; über den Beitrag einzelner Formanten wird nicht berichtet. Auch die Grundfrequenzen werden als sinkend beschrieben. Neben der geringen Sprecheranzahl ist die Auswahl dergestalt, dass die jüngste Person bei der frühen Aufnahme 29 Jahre alt war, die älteste bei der späten Aufnahme 88; es werden also Individualergebnisse aus völlig verschiedenen Lebensabschnitten beschrieben. Erstaunlich ist hier auch, dass der größte Abfall der mittleren Grundfrequenz über die 15 Jahre hinweg mit 6.6 Halbtönen (von 136 Hz auf 93 Hz) angegeben wird, was recht viel erscheint in dieser kurzen Zeit.

Decoster und Debruyne (2000) berichtet von mit dem Alter fallenden Grundfrequenzen bei männlichen Radioreportern, welche 30 Jahre später den gleichen Text nochmal lesen mussten. Harrington (2006) ist eine soziolinguistische Studie, die aber auch in Schwa-Vokalen Grundfrequenz und den ersten Formanten als mit dem Alter fallend feststellt. Harrington et al. (2007a) weitet dieses Vorgehen auf mehrere Sprecher aus (zwei britische Sprecherinnen, einen australischen und einen neuseeländischen Mann); gefunden werden ein Absinken von f0 und F1 bei allen vier Sprechern und Sprecherinnen, sowie von F2 bei einer der Frauen und bei beiden Männern. Schon dieses Paper schlug als Grund für die Verschiebungen in den Formanten auditorische Gründe vor: ausgehend von Traunmüller (1981, 1984, 1991a) wird der Abstand in Bark von *F1* und *f0* als eigentliches Maß für Vokalhöhe vorgeschlagen; wie Harrington et al. (2007a) feststellen, bleibt dieser Abstand allerdings nicht stabil, sondern verringert sich um 0,5 Bark, d.h. der Abfall von *F1* ist stärker als der von *f0*. Dies bedeutet, dass auch mit diesem Maß der Vokalraum sich etwas nach oben verschiebt. Ähnlich verhält es sich für die Daten in diesem Paper mit dem

F3-F2-Abstand in Bark, wie er von Syrdal und Gopal (1986) als Korrelat für das Feature [±*back*] vorgeschlagen worden war. Dieses Maß wird mit dem Alter etwas größer, oder, anders gedeutet, der Vokalraum verschiebt sich etwas nach hinten.

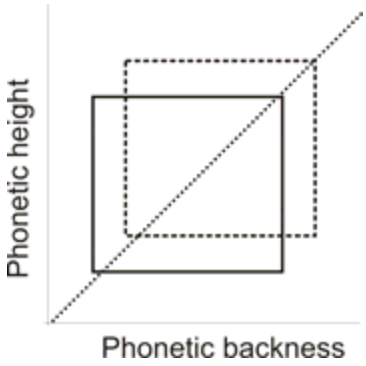

Fig. 5. *The proposed model of the shift in the speaker's space in the height x backness plane with increasing age (the dotted rectangle represents the same speaker's space at an older age).*

Abbildung 2.6: *Vokalraumverschiebungsmodell nach Harrington et al. (2007a, o.S.) nebst originaler Abbildungsunterschrift.*

Abbildung 2.6 zeigt das von Harrington et al. (2007a) vorgeschlagene Modell der Vokalraumverschiebung, welches sehr an die Abbildung 2.1 aus Watson und Munson (2007) erinnert, d.h. man sieht eine Verschiebung nach rechts oben entlang einer Diagonalen. Desweiteren erinnert dieses Modell sehr an die Abbildungen und Daten in Potter und Steinberg (1950) und Peterson (1961) oder auch in Turner und Patterson (2003), weshalb die Autoren davon ausgehen, dass diese Verschiebung entlang der Diagonalen im Vokalraum *nicht* in erster Linie als Vokalqualitätsunterschiede wahrgenommen werden, sondern vom Hörer genutzt werden, um unterschiedliche Sprecher und Sprecheralter voneinander zu unterscheiden, wobei, was die Vokalqualitäten angeht, für diese Information vom Hörer normalisiert wird, indem jeder sprecherspezifische Vokalraum entlang dieser Diagonalen angeordnet wird.

Wie ein weiteres Paper mit longitudinalen Daten aus der Gruppe um Harrington (Reubold et al., 2010) zeigte, indem es die Sprecherbasis auf 5 Sprecher der gleichen Varietät erweiterte und zwei dieser Sprecher zu mehr als nur zwei Zeitpunkten untersuchte, sind die altersbedingten Änderungen in *F2* und *F3 nicht* so systematisch, wie in Harrington et al. (2007a) angenommen. Bestätigung fand allerdings die Beobachtung einer Kovariation des ersten Formanten mit der Grundfrequenz. Die folgenden Untersuchungen lehnen sich stark an dieses Paper an, indem sie teilweise die gleichen Sprecher untersuchen, einige neue Sprecher hinzufügen, und einige alternative Hypothesen zu dieser Kovariation von *f0* und *F1*, die größtenteils bereits in Reubold et al. (2010) aufgestellt wurden, testen.

2.2 Veränderungen der Grundfrequenz und der Formanten in mehreren Sprechern – eine longitudinale Analyse

2.2.1 Teilexperiment 1: Wie beeinflusst das Alter die Grundfrequenz und die Formanten in Schwa-Vokalen?

Methode

Die Sprecher und ihre Sprachdaten Für diesen Teil der Arbeit wurden Schwa-Vokale in longitudinalen akustischen Daten von 5 Sprechern untersucht. Diese Sprecher waren die englische Königin Elisabeth die Zweite (⋆1926), die britische Schauspielerin Margaret Lockwood (⋆1916 †1990), der BBC Radiomoderator Roy Plomley (⋆1914 †1985), die frühere Britische Premierministerin Baroness Margaret Thatcher (⋆1925 †2013) sowie der britisch-amerikanische Journalist und Moderator Alistair Cooke (⋆1908 †2004). Die vier erstgenannten sprechen/sprachen eine Akzentvariante des SSB (Standard Southern British) genannten Standardenglischen, die sogenannte Received Pronunciation; bei Cooke machen sich neben der Received Pronunciation Einflüsse des General American genannten Akzentes des amerikanischen Englisch bemerkbar.

Für jeden der genannten Sprecher wurden zunächst Aufnahmen aus zwei Lebensabschnitten ausgewertet. Diese Aufnahmen entstammten den Archiven der Britisch Broadcasting Corporation (BBC); der zeitliche Abstand zwischen den „frühen“ und „späten“ Aufnahmen betrug zwischen 29 (bei Lockwood) und 35 (bei Thatcher) Jahren; das Sprecheralter betrug für die „frühen“ Aufnahmen zwischen 34 (bei Königin Elisabeth II) und 43 (bei Cooke) Jahren, bei den „späten“ variierte es zwischen 64 (bei Lockwood) und 73 (bei Cooke) Jahren. Die „frühen“ Aufnahmen von Cooke, Lockwood und Plomley stammen aus dem Jahr 1951, jene von Königin Elisabeth II und Margaret Thatcher aus dem Jahr 1960. Die „späten“ Sendungen wurden zwischen 1980 und 1995 aufgezeichnet. Weitere Details sind in Tabelle 2.1 zu finden.

Bei den Aufnahmen der Königin handelt es sich um zwei ihrer jährlich ausgestrahlten Weihnachtsansprachen (Harrington, 2006; Harrington et al., 2000a, 2007b). Beide Aufnahmen Margaret Lockwoods stammen von ihren Gastauftritten in der Sendung „Desert Island Discs“; die Aufnahmen unterscheiden sich vom Sprechstil dahingehend ein wenig, als die erste Aufnahme offensichtlich einstudierte Sprache enthält, die „späte“ Aufnahme hingegen Spontansprache. Roy Plomley war der Moderator der Sendung „Desert Island Discs“ doch stammen die „frühe“ und „späte“ Aufnahme aus anderen Sendungen als aus der mit Margaret Lockwood (für die beiden Lockwood-Auftritte in „Desert Island Discs“ ist nur sehr wenig Sprachmaterial von Plomley verfügbar). Beide Thatcher-Aufnahmen enthalten Interviewmaterial von in etwa gleichem Sprachstil ruhiger Konversation. Bei Cookes Sprachmaterial handelt es sich nicht um Spontansprache; es ist nicht leicht zu entscheiden, ob es sich um Lesesprache oder um vorher nach einem Manuskript einstudierte Sprache handelt; jedenfalls blieb dem Höreindruck nach der Sprachstil über die Jahrzehnte hinweg

SprecherIn	früh (Alter)	spät (Alter)	Altersdifferenz
Königin Elisabeth II	1960 (34)	1994 (68)	34
Margaret Lockwood	1951 (35)	1980 (64)	29
Margaret Thatcher	1960 (35)	1995 (70)	35
Alistair Cooke	1951 (43)	1981 (73)	30
Roy Plomley	1951 (37)	1985 (71)	34

Tabelle 2.1: Die fünf Sprecher, Sendungsjahr und Alter (in Klammern) für die frühen und späten Aufnahmen. Die letzte Spalte zeigt die Altersdifferenz zwischen früher und später Aufnahme.

vergleichbar.

Die Datenmenge pro Sprecher variiert nicht unerheblich zwischen den Sprechern: während für Cooke jeweils fast viertelstündige Aufnahmen zur Verfügung standen (13 Minuten, 24 Sekunden (früh) bzw. 13 Minuten 20 Sekunden (spät)), waren es für die Königin nur jeweils ca. 5 Minuten (4 Minuten 36 Sekunden (früh) gegenüber 5 Minuten 32 Sekunden (spät)); für Lockwood steht eine 5 Minuten 30 Sekunden lange frühe Aufnahme einer 12 Minuten langen späten Aufnahme gegenüber; sehr kurz sind die Netto-Dauern des Materials bei Plomley (53 Sekunden (früh), 2 Minuten 17 Sekunden (spät)) und Thatcher (47 Sekunden (früh), eine Minute 16 Sekunden (spät)).

Vorverarbeitung der Daten Da das Interesse in dieser Studie den langfristigen nichtphonetischen Effekten des Alters auf die Grundfrequenz und die Formantfrequenzen galt, wurde der Schwa als Untersuchungsgegenstand gewählt, da für diesen Laut ein Wandel weder bekannt noch wahrscheinlich ist. Harrington (2006) folgend wurden nur solche Schwa-Laute gewählt, welche in mehrsilbigen Inhaltswörtern auftreten; ausgeschlossen wurden darunter diejenigen reduzierten Vokale, welche zumindest bei sorgfältiger Artikulation mit einem /ɪ/ produziert werden können und auch so in den meisten Lexika, wie z. B. dem Cambridge English Pronouncing Dictionary (D. Jones & Roach, 2009) als Aussprachenorm transkribiert werden (z. B. in dem Wort *roses* (*Rosen*); im Gegensatz dazu würde der Schwa in *Rosa's* (Genitiv von *Rosa*) mit einbezogen werden, obschon *roses* und *Rosa's* homophon produziert werden können). Da die Audiodaten zu unterschiedlichen Zeitpunkten und unter unterschiedlichen Bedingungen von den Archiven der BBC zur Verfügung gestellt wurden, unterscheiden sich die Abtastfrequenzen: 16kHz (Königin Elisabeth II und Lockwood), 22 kHz (Cooke) und 24 kHz (Plomley und Thatcher). Die Grundfrequenz sowie die Formantfrequenzen wurden mit den Standardanwendungen von tkassp 2.0 (*f0ana* (Schaefer-Vincent, 1983) für die Grundfrequenzschätzung und *forest* für die Formantberechnung) bei einer Fensterlänge von 30 ms und einer Schrittbreite von 5 ms berechnet. Die Daten wurden

nicht manuell nachkorrigiert, Schwas mit Werten unterhalb eines Schwellwertes von 40 Hz in mindestens einem der gemessenen Parameter jedoch von der Analyse ausgeschlossen, so dass sich die in Tabelle 2.2 zu findende Anzahl analysierter Schwas ergibt. Insgesamt wurden 447 „frühe“ mit 617 „späten“ Schwas verglichen. Wegen der Unterschiede an zur Verfügung stehendem Material pro Sprecher ergeben sich auch erhebliche Unterschiede an Anzahl von Schwas pro Sprecher (siehe Tabelle 2.2). Für jeden dieser Schwas wurde ein Wert für f0 und F1, F2 und F3 am zeitlichen Mittelpunkt extrahiert.

Alle nachfolgenden Analysen wurden mit dem package Emu/R (Harrington & IPS LMU Muenchen & IPDS CAU Kiel, 2011) in *R* (R Development Core Team, 2010), die anschließenden statistischen Analysen mittels der *R*-Pakete car (Fox & Weisberg, 2011) bzw. lme4 (Bates, Maechler & Bolker, 2011) durchgeführt.

SprecherIn	**früh**	**spät**	**Gesamt**
Königin Elisabeth II	204	175	379
Margaret Lockwood	74	137	211
Margaret Thatcher	30	118	148
Alistair Cooke	110	140	250
Roy Plomley	29	47	76
Gesamt	447	617	**1064**

Tabelle 2.2: Die Anzahl der /ə/-Vokale für fünf Sprecher für die frühen und späten Aufnahmen

Ergebnisse

Medianwerte und Verteilung der Daten in den Boxplots der Abbildung 2.7 zeigen, das tendenziell f0 und F1 bei „spät“ niedriger sind; die einzigen möglichen Ausnahmen sind die Grundfrequenz bei Plomley und der erste Formant bei Lockwood [8], da diese bei dem späten Aufnahmezeitpunkt nur geringfügig tiefer liegen. Für F2 lässt sich keine eindeutige Tendenz feststellen, und für F3 eine leichte Tendenz für höhere Werte bei spät. Für alle vier Parameter waren die Voraussetzungen für eine *ANOVA mit Messwiederholung* gegeben (*F-Tests*: f0: $F[4,4] = 1,32$; n.s., $F1 : F[4,4] = 0,8$; n.s., F2: $F[4,4] = 1,27$; n.s., F3: $F[4,4] = 0,48$; n.s.; *Shapiro-Tests*: f0 (früh):$W = 0,94$; n.s., f0(spät):$W = 0,86$; n.s., F1 (früh):$W = 0,91$; n.s., F1 (spät):$W = 0,9$; n.s., F2 (früh): $W = 0,96$; n.s., F2 (spät): $W = 0,82$; n.s., F3 (früh): $W = 0,94$; n.s. F3 (spät): $W = 0,92$; n.s.). Mittels *ANOVA mit Messwiederholung*

[8] Bei Lockwood fällt generell die für eine Frauenstimme ungewöhnlich tiefe Lage der Messwerte auf

(repeated-measures ANOVA oder *RM-ANOVA)* für jeden der Parameter f0, F1, F2 oder F3 als abhängiger Variable, *Altersstufe* als unabhängiger Variable (ein zweistufiger Faktor: früh/ spät) und *Sprecher* als Random-Faktor ergaben sich signifikante *Alters*-Effekte sowohl für f0 ($F[1,4] = 26,2; p < 0,01$) als auch für F1 ($F[1,4] = 10,7; p < 0,05$), nicht jedoch für F2 ($F[1,4] = 0,0$; n.s.), und ebensowenig für F3 ($F[1,4] = 0,7$; n.s.); siehe auch Abbildung 2.7. Um zusätzlich den Effekt des Geschlechtes zu untersuchen, wurden die Wer-

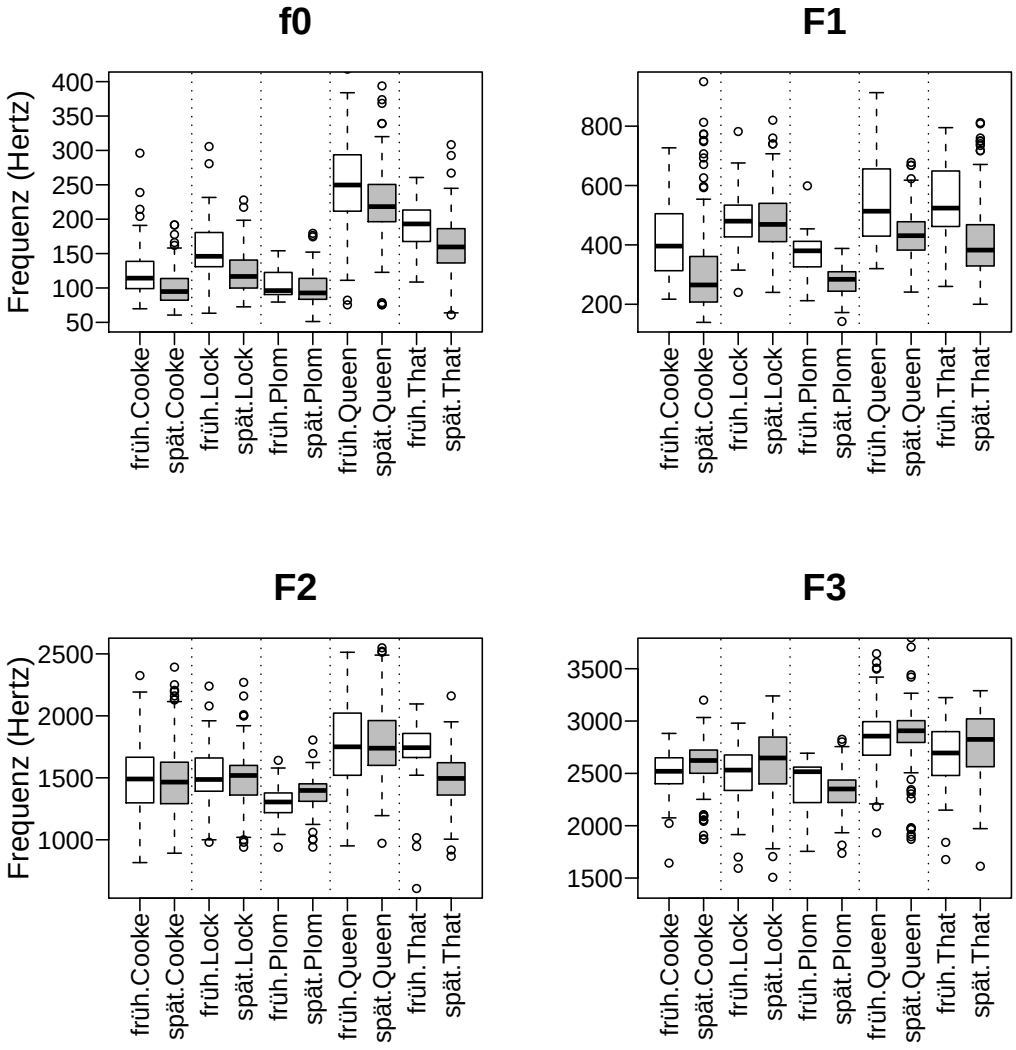

Abbildung 2.7: *Verteilung von f0, sowie von F1-3 (in Hertz) in den Schwas von fünf Sprechern (Cooke, Lockwood, Plomley, Königin Elisabeth II, Thatcher) zu jeweils zwei Aufnahmezeitpunkten*

te von f0 und der ersten drei Formanten mit der in Emu/R implementierten Formel aus Traunmüller (1990a) nach Bark konvertiert, um geschlechtsspezifische lagebedingte Varia-

tion der Unterschiede zwischen *früh* und *spät* so weit als möglich auszuschließen; wegen der geringen Sprecheranzahl pro *Geschlecht* wurden anstelle von ANOVAs mit Messwiederholung *Lineare Gemischte Modelle (Linear mixed models)* berechnet, mit f0, F1, F2 oder F3 als abhängiger Variable, den Zwischen-Subjekt-Faktoren $Altersstufe$ ($früh$ vs. *spät*) und *Geschlecht* (*männlich* und *weiblich*) als unabhängiger Variable und *Sprecher* als Zufallsvariable. Das generelle Bild verändert sich hierdurch kaum (siehe Abbildung 2.8): für F2 gibt es nach wie vor keine signifikanten Alterseffekte ($F[1, 60] = 0,05$; n.s.) und nur ein schwacher Geschlechtseffekt ($F[1, 60] = 4,3; p < 0,05$); F3 weist auch bei dieser Analyse keine signifikante Beeinflussung durch *Geschlecht* ($F[1, 60] = 0,13; p = 0,72$; n.s.) oder *Alter* ($F[1, 60] = 1,2; p = 0,28$; n.s.) auf. Der erste Formant ist hochsignifikant sowohl von *Geschlecht* ($F[1, 60] = 300,8; p < 0,01$) als auch von *Alter* ($F[1, 60] = 148,4; p < 0,01$) beeinflusst, und es ist keine Interaktion *Geschlecht* × *Alter* festzustellen. Eine leichte Interaktion ($F[1, 60] = 4,7; p < 0,05$) dieser Art findet man hingegen bei der Grundfrequenz, welche erstaunlicherweise nur schwach von *Geschlecht* ($F[1, 60] = 6,2; p < 0,05$) [1] , aber hochsignifikant von *Alter* abhängt ($F[1, 60] = 102,8; p < 0,01$). Getrennt nach Geschlechtern ausgeführt ergeben Lineare Gemischte Modelle jedoch für beide Geschlechter hochsignifikanten Einfluss des $Alters$ (Männer: $F[1, 60] = 20,8; p < 0,001$, Frauen: $F[1, 60] = 55,3; p < 0,001$). Wie der Interaktions-Plot in Abbildung 2.9 zeigt, ist die Abhängigkeit der Grundfrequenz von *Alter* bei den weiblichen Sprechern jedoch deutlicher ausgeprägt; jedoch sinkt bei beiden Geschlechtergruppen f0 mit dem Alter.

Diskussion der Ergebnisse

Die Beschränkung auf segmentierte, vergleichbare Materialien zu zwei Zeitpunkten, die circa dreißig Jahre auseinanderliegen, beschränkt nicht unerheblich die Auswahlmöglichkeiten. Daher ist dies eine der relativ wenigen Studien (neben Brückl (2007); Decoster und Debruyne (1997, 2000); Harrington et al. (2007a); Reubold et al. (2010); Russell et al. (1995); Verdonck-De Leeuw und Mahieu (2004); hierbei Reubold et al. (2010) mit den gleichen Sprechern wie in dieser Studie), die erwachsene Sprecher akustisch in einer Längsschnittstudie untersuchen. Die Resultate dieses Experiments sind bezüglich f0 konsistent mit denen aus Querschnitts- (Baken, 2005; Linville, 1996; Nishio & Niimi, 2008) und Längsschnittstudien (Decoster & Debruyne, 2000; Harrington et al., 2007a), diejenigen bezüglich der Altersabhängigkeit vornehmlich des ersten Formanten (mit einem altersbedingten Absinken von F1) sind konsistent mit den Querschnittsstudien in (Linville & Fisher, 1985a; Xue & Hao, 2003).

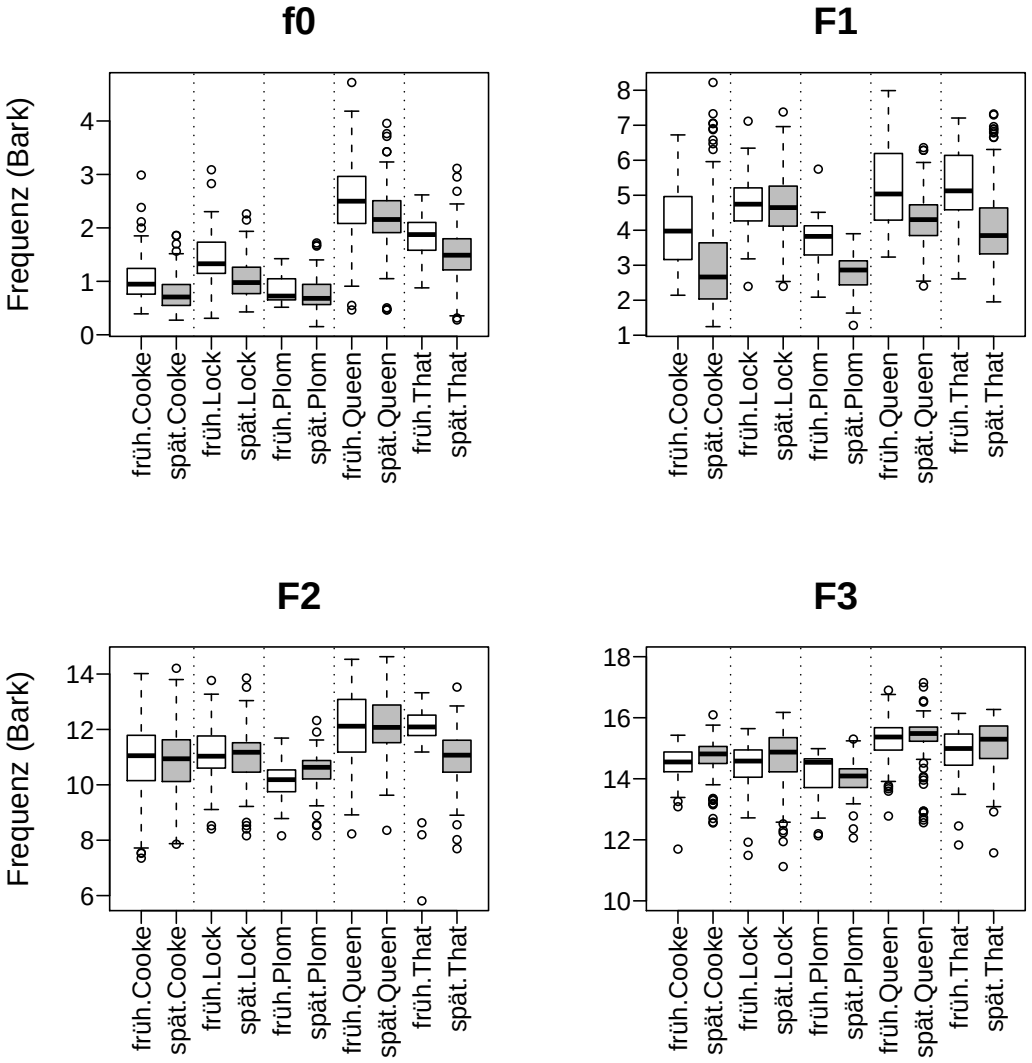

Abbildung 2.8: *Verteilung von f0, sowie von F1-3 (in Bark) in den Schwas von fünf Sprechern (Cooke, Lockwood, Plomley, Königin Elisabeth II, Thatcher) zu jeweils zwei Aufnahmezeitpunkten*

2.2.2 Teilexperiment 2: Sind Messungen anhand stimmhafter Frames vergleichbar mit Messungen in Schwa-Vokalen?

Methode

Vorüberlegungen Um mehr Material als die Daten der zwei Männer und drei Frauen zu jeweils zwei Zeitpunkten wie in Experiment 2.2.1 untersuchen zu können, wird es nötig sein, auf zeit-, arbeits- und damit kostenintensive Segmentationen zu verzichten und Mit-

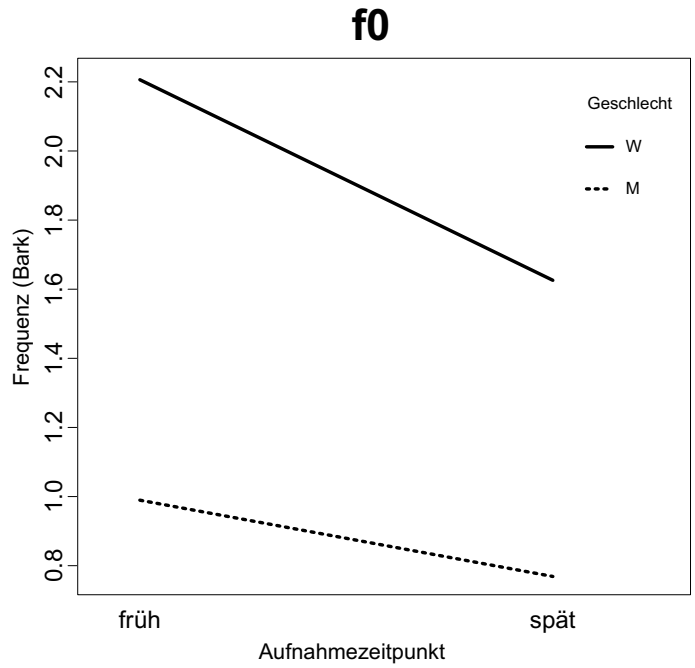

Abbildung 2.9: *Interaktion der mittleren Grundfrequenz, gemessen in Schwa-Vokalen, zwischen männlichen ($N = 2$) und weiblichen ($N = 3$) Sprechern*

telwerte für das gesamte Sprachmaterial eines Sprechers zu einem bestimmten Zeitpunkt heranzuziehen. Im vorliegenden Fall wird dies bedeuten, für jeden Sprecher zu jedem gegebenen Zeitpunkt jeweils einen Wert für f0 und F1-3 zu erhalten, der sich als arithmetisches Mittel aus Messungen innerhalb aller stimmhaften Signalanteile zu dem gegebenen Zeitpunkt ergibt, d.h. Werte werden in jedem Messfenster mit stimmhafter Anregung gemessen und danach gemittelt. Hierbei stellt sich jedoch die Frage, inwiefern diese Methode mit den Messungen aus Experiment 2.2.1, also mit der Mittelung anhand jeweils eines Wertes pro Schwa-Vokal vergleichbar ist. Für die Verwendung des Schwa-Vokals sprach, dass dieser Vokal als Neutralvokal wohl am ehesten die Eigenschaften des Ansatzrohres und seiner Veränderung über die Zeit wiedergibt, ohne dass vom (englischen) Schwa bekannt wäre, dass er von Lautwandel oder anderen soziophonetischen Prozessen beeinflusst wäre; dies bedeutet, dass der Autor als Prämisse davon ausgeht, dass der Schwa von allen Vokalen am wenigsten von Veränderungen der Artikulation beeinflusst sein sollte, gleichgültig, was die Quelle solcher Veränderungen sein sollte. Andererseits wird von Messungen anhand aller stimmhaften Sprachsignalanteile ebenfalls erwartet, dass ihre Mittelung zu vergleichbaren Ergebnissen führen sollte, solange der mögliche Vokalraum nicht assymetrisch ausgenutzt wird. Der Hintergedanke hierbei ist, dass die Mittelung aller möglichen Formantwerte zu in etwa den gleichen Werten führen sollte, die man als Formantwerte für eine neutrale Zungenlage (also für den Schwa-Vokal) erwarten sollte.

Die Sprachdaten Um die Vergleichbarkeit von Messungen anhand von Schwa-Daten mit Messungen anhand stimmhafter Sprachsignalanteile zu ermitteln, wurde das gleiche Sprachmaterial verwendet wie in Experiment 2.2.1; die anhand der Schwa-Segmentation ermittelten Werte (je ein Wert pro Schwa, gemessen am zeitlichen Mittelpunkt) für Schwa-Vokale wurden pro Sprecher gemittelt, so dass für jeden der fünf Sprecher (Cooke, Plomley, Lockwood, Königin Elisabeth II, Thatcher) je ein Wert für f0, F1, F2 und F3 ermittelt wurde. Daneben wurden f0 und Formanten auch in allen Signalfenstern, die stimmhafte Anregung aufwiesen, gemessen, Werte aus Signalfenstern, die in einem der drei Formantwerte Nullstellen aufwiesen, ausgeschlossen, und die Werte anschließend pro Sprecher gemittelt, so dass auch bei dieser Methode je ein Wert pro Sprecher für f0 und die Formanten F1-3 ermittelt wurde. Diese parallelen Vektoren von Werten wurden jeweils mittels eines gepaarten t-tests miteinander verglichen.

Ergebnisse

Die gepaarten t-tests ergaben weder für f0 ($t[4] = 0,18$; n.s.) noch für die ersten drei Formanten (F1: $t[4] = 0,81$; n.s.; F2: $t[4] = 1,22$; n.s.; F3: $t[4] = 1,46$; n.s.) signifikante Unterschiede zwischen gemittelten Messwerten aus Schwa-Vokalen bzw. stimmhaften Sprachsignalanteilen.

Diskussion der Ergebnisse

Dieser höchst einfache Test diente dazu, die vermutete Vergleichbarkeit von Messungen anhand des zeitlichen Mittelpunktes von Schwa-Vokalen und von Messungen anhand stimmhafter Signalanteile des gesamten Sprachsignals zu überprüfen. Da keine signifikanten Unterschiede festgestellt werden konnten, wird der Autor dazu übergehen, in den folgenden Experimenten Messwerte aus stimmhaften Sprachsignalanteile zu verwenden. Während eine Voraussage getroffen wurde, dass die Mittelung der Formantwerte aus allen möglichen Kontexten zu ähnlichen Werten führen sollte, als wenn man in Schwa-Vokalen zum zeitlichen Mittelpunkt misst, war das Verhältnis der Grundfrequenz aus beiden Messmethoden nicht vorhersagbar. Am ehesten hätte man spekulieren können, dass Grundfrequenzen in Schwa-Vokalen tiefer sein sollten als der Mittelwert der gesamten Äußerung, da Schwa-Vokale (im hier untersuchten Englischen) ausschließlich in unbetonten Silben auftreten. Stattdessen ergeben allerdings die vorliegenden Tests auch keine signifikanten Unterschiede zwischen mittleren f0-Werten aus Schwa-Daten bzw. stimmhaften Signalanteilen.

2.2.3 Teilexperiment 3: Beeinflusst das Alter die mittlere Grundfrequenz und die Mittelwerte der Formanten 1-3 in stimmhaften Signalabschnitten?

Methode

Die Sprecher und ihre Sprachdaten Als erste Anwendung der Methode, die Altersabhängigkeit der Grundfrequenz- und Formantwerte anhand der Messwerte aus stimmhaften Signalabschnitten zu bestimmen, wurde eine dem ersten Experiment vergleichbare Vorgehensweise gewählt, d.h. bei mehreren Sprechern wurden Daten zu zwei Zeitpunkten in ihrem Leben, die mindestens 20 Jahre auseinander lagen, analysiert. Auch hierbei ist selbstverständlich darauf zu achten, dass die Aufnahmebedingungen und der Sprechstil vergleichbar sind, was bedauerlicherweise die Suche nach geeigneten Daten erschwert. So konnten im vorliegenden Fall nur drei männliche Sprecher den fünf Sprechern aus Experiment I hinzugefügt werden, so dass schlussendlich die Daten von 5 Männern und 3 Frauen zur Verfügung standen. Bei den zusätzlichen Sprechern handelt es sich um den hauptsächlich durch Tier- und Naturdokumentationen bekanntgewordenen Funk- und Fernsehmoderator David Attenborough (⋆1926), den Satiriker und Humoristen Frank Muir (⋆1920 †1998) und um den Rundfunkreporter und später in leitende Funktionen der BBC aufgestiegenen Frank Gillard (⋆1909 †1998). In allen drei Fällen handelt es sich sowohl bei der frühen wie der späten Aufnahme um Interviews, bei denen die Sprecher zu ihrem beruflichen Leben befragt wurden, d.h. bei allen drei Sprechern ist vergleichbares Material zu beiden Aufnahmezeitpunkten vorhanden (Attenborough: Details zu seinem Leben als Tierfilmer; Muir: sein Berufsleben als Humorist in der BBC; Gillard: seine administrativen Funktionen innerhalb der Leitung der BBC). Die Aufnahmen stammen aus den Jahren 1956 / 1976 (Attenborough), 1969 /1997 (Muir), sowie 1964 / 1987 (Gillard), d.h. die Zwischenräume zwischen früh und spät sind nicht unerheblich kürzer als bei den anderen Sprechern. Details über das pro Sprecher und Aufnahmezeitpunkt vorhandene Material sowie das Sprecheralter sind in Tabelle 2.3 zu finden. Das Alter für die früh-Aufnahme schwankte zwischen 30 und 55 Jahren, das für die spät-Aufnahme zwischen 50 und 78; die Altersdifferenz zwischen beiden Zeitpunkte schwankte zwischen 20 und 35 Jahren.

Wie in Experiment I wurde auch für das vorliegende Experiment das Sprachmaterial für alle 8 Sprecher in Ausschnitte von maximal einer Minute Länge geschnitten, sofern (bei den Interviews) der Äußerungsturn nicht durch den Interviewer unterbrochen wurde, so dass in Ausnahmefällen auch Ausschnitte kürzer als eine Minute vorkommen konnten. Für jeden der so entstandenen Aufnahmeausschnitte wurde durch Analyse aller stimmhaften Signalabschnitte mit den gleichen Einstellungen für *f0ana* und *forest* wie in Experiment I (Fensterlänge: 30 ms; Schrittbreite: 5 ms) sowie durch anschließende Berechnung arithmetischer Mittelwerte je ein Wert für f0 und die ersten drei Formanten ermittelt. Diese Werte wurden zu Bark-Werten umgerechnet, um die geschlechtsbedingten Lageunterschiede zu berücksichtigen und damit die natürlich für weibliche Sprecher größeren Unterschiede im Hertz-Bereich zwischen *früh* und *spät* auszugleichen.

SprecherIn	früh (Alter)	spät (Alter)	Alters-differenz	Minuten, Sekunden (früh)	Minuten, Sekunden (spät)
Attenborough	1956 (30)	1976 (50)	20	14'26	10'46
Cooke	1951 (43)	1981 (73)	30	13'27	26'48
Gillard	1964 (55)	1987 (78)	23	3'58	23'03
Muir	1969 (49)	1997 (77)	28	3'57	15'30
Plomley	1951 (37)	1985 (71)	34	1'44	3'07
Lockwood	1951 (35)	1980 (64)	29	5'30	11'59
Elisabeth II	1960 (34)	1994 (68)	34	4'37	5'42
Thatcher	1960 (35)	1995 (70)	35	2'28	15'25

Tabelle 2.3: Übersicht über die Aufnahmezeitpunkte, das Alter des jeweiligen Sprechers zu diesen Zeitpunkten, die Altersdifferenz sowie die pro Sprecher und Aufnahmezeitpunkt verfügbare Menge an Sprechdaten in Minuten und Sekunden

Ergebnisse

Da es zusammen für alle Sprecher 349 Aufnahmeausschnitte (und somit die gleiche Anzahl an Messwerten pro Parameter) gab, und diese Zahl bereits im unteren Bereich der notwendigen Mindestanzahl für Eingabedaten für die abhängige Variable bei Linearen Gemischten Modellen liegt, wurde beschlossen, den konservativeren Weg zu beschreiten und die Daten mittels der Methode *ANOVA mit Messwiederholung* zu analysieren. Die Voraussetzungen hierfür (dass die Varianzen der Stufen eines Faktors voneinander nicht signifikant unterschiedlich sind und dass die Verteilung der Werte innerhalb der Stufen nicht signifikant von einer Normalverteilung abweicht) wurde mittels $F-Tests$ und *Shapiro-Wilk normality tests* untersucht und ergab keine Abweichung von den Voraussetzungen ($F-Tests$: f0: $F[7,7] = 1,8$; n.s., F1: $F[7,7] = 1,1$; n.s., F2: $F[7,7] = 0,6$; n.s., F3: $F[7,7] = 0,6$; n.s. ; *Shapiro-Wilk*: f0 (früh): $W = 0,9$; n.s., f0 (spät): $W = 0,8$; n.s., F1 (früh): $W = 0,9$; n.s., F1 (spät): $W = 0,9$; n.s. , F2 (früh): $W = 0,8$; n.s., F2 (spät): $W = 0,9$; n.s., F3 (früh): $W = 0,9$; n.s., F3 (spät): $W = 1,0$; n.s.). Für die 4 *ANOVAs mit Messwiederholung* war entweder $f0$, $F1$, $F2$ oder $F3$ die abhängige Variable; Zwischensubjektfaktor war das *Geschlecht* (m/w), der *Aufnahmezeitpunkt* (*früh* vs. *spät*) war Innersubjektfaktor; die Variation durch die Sprecher wurde ausgeklammert.

Für den ersten Formanten ergab sich ein hochsignifikanter Effekt für *Alter* ($F[1,6] = 54,6; p < 0,01$), sowie ein *Geschlechts*effekt ($F[1,6] = 29,5; p < 0,01$), und keine Interak-

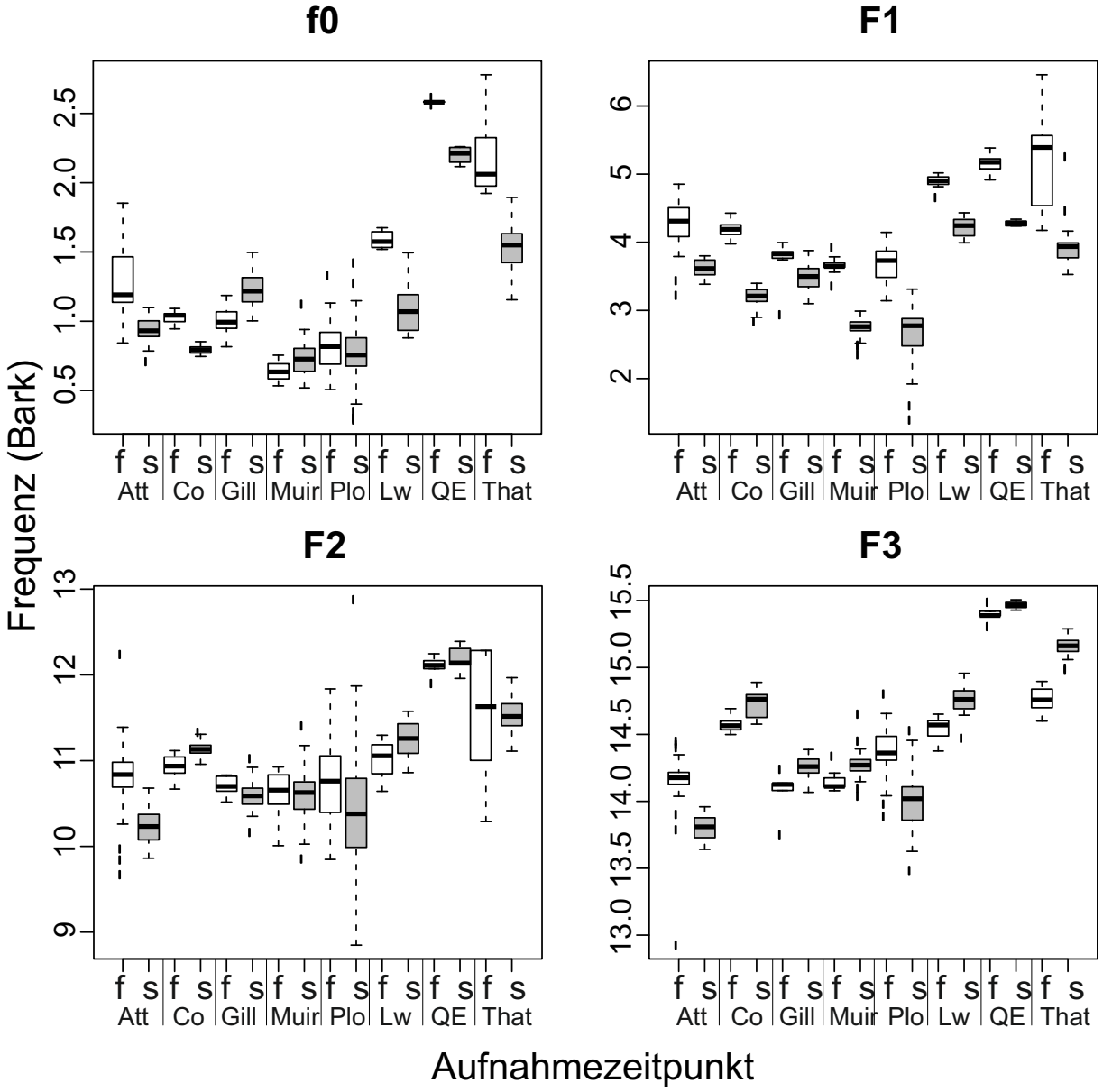

Abbildung 2.10: *Verteilung von f0, sowie von F1-3 (in Bark) in den stimmhaften Signalanteilen von acht Sprechern (Attenborough (Att), Cooke (Co),Gillard (Gill), Muir (Muir), Plomley (Plo), Lockwood (Lw), Königin Elisabeth II (QE), Thatcher (That)) zu jeweils zwei Aufnahmezeitpunkten (f (früh) – s(spät))*

tion ($F[1,6] = 0,5$; n.s.) zwischen beiden Faktoren. Für die Formanten 2 und 3 ergaben sich zwar - wie zu erwarten war - leichte Effekte für *Geschlecht* ($F2 : F[1,6] = 13,7; p < 0,05$; F3: $F[1,6] = 12,2; p < 0,05$), aber keinerlei Effekte für *Alter* (F2: $F[1,6] = 0,37$; n.s.; F3: $F[1,6] = 0,33$; n.s.) und auch keine Interaktionen *Geschlecht* × *Alter* (F2: $F[1,6] = 2,0$; n.s.; F3: $F[1,6] = 2,4$; n.s.) , d.h. es war für Formanten jenseits des ersten keinerlei Ten-

denz für einen Einfluss des Alters festzustellen. Für die Grundfrequenz ergab sich (trivialerweise) ein Effekt für *Geschlecht* ($F[1,6] = 14,0; p < 0,01$), aber auch ein Effekt für *Alter* ($F[1,6] = 10,57; p < 0,05$) und eine Interaktion zwischen beiden Faktoren ($F[1,6] = 8,67; p < 0,05$). Bei post-hoc durchgeführten gepaarten t-Tests mit $Bonferroni$-$Korrektur$ (Bonferroni-Faktor=6) ergab sich für keines der Geschlechter ein signifikanter Effekt des Alters (w: $t[2] = 7,1$; n.s., m: $t[4] = 0,7$; n.s.) . Bei Anwendung von Jonathan Harringtons Tukey.rm-Funktion[9] zur Durchführung von post-hoc-Tests (mit den Faktoren *Geschlecht* und *Alter*) ergab sich jedoch ein Alterseffekt für Frauen ($p < 0,05$), jedoch kein Effekt des *Alters* für Männer. Dies ist auch aus dem Interaktions-Plot ersichtlich: für die Frauen sinkt die Grundfrequenz deutlich mit dem Alter, für Männer bleibt sie im Durchschnitt praktisch gleich.

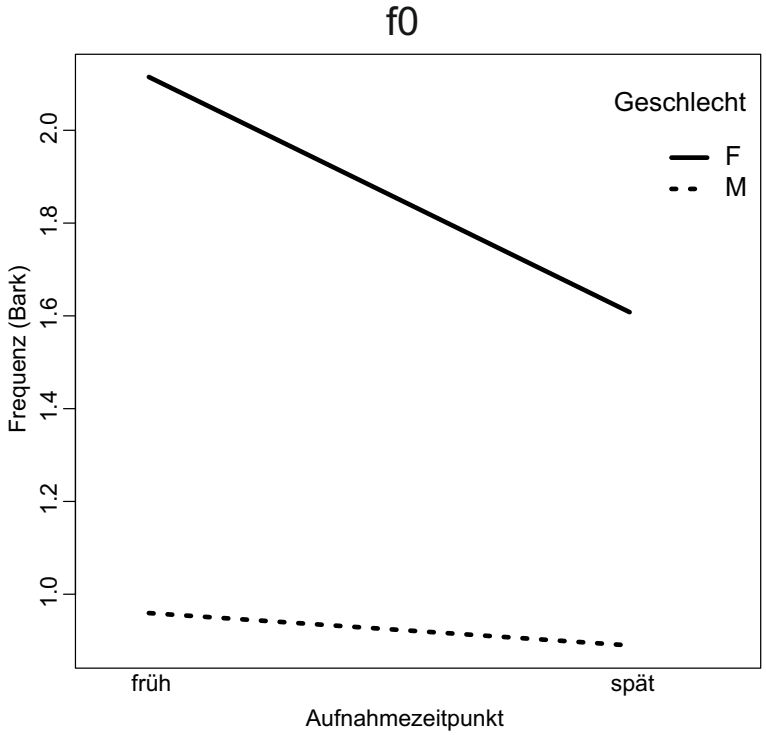

Abbildung 2.11: *Interaktion der mittleren Grundfrequenz, gemessen in stimmhaften Signalabschnitten, zwischen männlichen ($N = 5$) und weiblichen ($N = 3$) Sprechern*

Dies ist erklärlich aus dem inkonsistenten Verhalten der Grundfrequenz (siehe Abbildung 2.11, linkes oberes Viertel) bei Männern: während f0 bei Attenborough und Cooke deutlich (und eventuell leicht bei Plomley) sinkt, steigt sie mit zunehmendem Alter bei Muir und Gillard.

[9] http://www.phonetik.uni-muenchen.de/%7Ejmh/lehre/sem/ss08/statinR08/anova1

Diskussion der Ergebnisse

Die Analyse der hier um drei Männer gegenüber Teilexperiment I erweiterte Sprechergruppe bestätigt bezüglich der Formanten 1-3 das Bild, das Messungen anhand von Schwa-Vokalen gezeigt hatten, auch für Messungen anhand stimmhafter Signalabschnitte. Sowohl F2 und F3 zeigen sich vom Alter der Sprecher nicht signifikant und über die Sprecher hinweg vor allem nicht konsistent beeinflusst, d.h. Alter scheint kein bestimmender Faktor für die Lage höherer Formanten zu sein. Allerdings hat Alter offenbar große Effekte auf das arithmetische Mittel des ersten Formanten, denn dieser Wert fällt konsistent mit dem Alter. Dies ist konsistent mit den Befunden in der Literatur (Linville, 1987b; Linville & Fisher, 1985a, 1985b; Linville & Rens, 2001; Xue & Hao, 2003). Ein anderes Muster ist für die Grundfrequenz festzustellen. Wie ein Vergleich der Ergebnisse für die gleichen drei Frauen in der Interaktions-Plots in den Abbildungen 3 und 5 aufzeigt, ergeben sich in der Tat sehr ähnliche Ergebnisse für Messungen einerseits anhand von Schwa-Vokalen und andererseits anhand von stimmhaften Signalabschnitten, so dass beide Methoden wohl als gleichwertig gelten dürfen. Es ergibt sich bei beiden Methoden, dass die Grundfrequenz mit dem Alter fällt, und zwar aus einer vergleichbaren Lage bei den frühen Schwas und stimmhaften Signalabschnitten in ebenfalls vergleichbare Lage bei den späten Aufnahmen. Die Hinzunahme von drei weiteren Männern hat aber das in den Schwa-Vokalen als recht eindeutig erscheinende Ergebnis, das f0 auch für Männer fällt, verwirrt, da festzustellen ist, dass zwar bei zwei, möglicherweise sogar drei Sprechern (Attenborough, Cooke, und möglicherweise Plomley) die Grundfrequenz tatsächlich fällt, bei zwei Männern jedoch ansteigt. Nun könnte man annehmen, dass dies für Männer bezüglich der Grundfrequenz bedeutet, dass ebenso wie für die höheren Formanten kein Alterseffekt vorliegt. Befunde aus der Literatur hingegen lassen jedoch vermuten, dass hier eher eine methodische Schwäche vorliegt: bei beiden Sprechern (Muir und Gillard), bei denen die Grundfrequenz ansteigt, werden bereits etwas fortgeschrittenere Lebensabschnitte (5tes bis 8tes Lebensjahrzehnt) untersucht als bei Attenborough (3test bis 5tes Lebensjahrzehnt) oder bei Cooke und Plomley (4tes bis 7tes/6tes Lebensjahrzehnt), wobei bei Plomley das altersbedingte f0-Absinken nur mehr schwach (wenn überhaupt) ausgeprägt ist. Linville beschreibt auf den Seite 172/173 ihres Buches (Linville, 2001), dass die mittleren Grundfrequenzen bei männlichen Populationen zunächst vom Alter von 20 bis 40/50 Jahren absinken, während sie nach dem vierten Lebensjahrzehnt wieder ansteigen. Da es sich bei diesen Daten um eine Kompilation mehrerer Studien handelt, ist es nicht eindeutig, wo genau der Umkipppunkt zu bestimmen ist und wie stark die interindividuelle Variation zwischen männlichen Sprechern einzuschätzen ist. Baken (2005) beschreibt sogar einen früheren Umkipppunkt vor dem 40sten Lebensjahr. Weitere Studien und Studienzusammenfassungen lassen ähnliche Verläufe vermuten (Brown et al., 1991; Hollien & Shipp, 1972); siehe auch die Abbildung 2.3 in der Einleitung. Die hier vorliegenden Daten könnten also dahingehend gedeutet werden, dass die Grundfrequenz bei Attenborough und Cooke noch im Sinken begriffen ist, während für Muir und Gillard ein Analyseausschnitt gewählt wurde, in dem bei beiden Sprechern f0 bereits wieder ansteigt. Plomleys Daten könnten so gedeutet werden, dass sein Umkipppunkt bereits überschritten wurde und daher als Zufallsbefund die mittleren Grundfrequenzen zu

den beobachteten Zeitpunkten sich nur wenig unterscheiden. Dies ist allerdings eine rein spekulative Deutung, die sich zwar auf Daten aus der Literatur stützt, jedoch genauere Untersuchung erfordert. Eine solche Untersuchung muss zum Ziel haben, Sprecher zu mehr als nur zwei Zeitpunkten zu untersuchen, um den Verlauf der Grundfrequenz, aber auch der Formanten nachzuverfolgen, um mögliche Nichtlinearitäten aufzudecken. Diese Aufgabe wird vom nächsten Experiment angegangen.

2.2.4 f0- und F1-Veränderungen über mehrere Jahrzehnte in zwei Sprechern

Bedingungen für geeignete Korpora

Ziel dieses Experiments ist es, vergleichbare gesprochene Sprache einer Sprecherin und eines Sprechers über einen Zeitraum von mehreren Jahrzehnten in regelmäßigen Abständen akustisch zu untersuchen, um eventuell auftretende Non-Linearitäten verfolgen zu können. Dieses Ziel wird, nachdem in dieser Arbeit festgestellt wurde, dass Messungen in stimmhaften Signalabschnitten und Messungen in Schwa-Vokalen vergleichbare Ergebnisse liefern, nur noch vor eine Herausforderung gestellt, da zeit- und kostenintensive Segmentierung und Etikettierung entfallen können, nämlich das Finden geeigneten Materials. Zur Untersuchung von Lautwandel innerhalb einer Person hat Harrington in mehreren Studien (Harrington et al., 2000a; Harrington, 2006; Harrington et al., 2007a, 2007b) als Datengrundlage die Weihnachtsansprachen von Königin Elisabeth II verwendet (siehe Office of Public Sector Information - Information Policy Team (2011)), die wahrscheinlich einzigartig als Testmaterial für eine Longitudinalstudie sind. Erstens sind diese Ansprachen deswegen nahezu ideal, da sie im jährlichen Rhythmus aufgenommen wurden (nur einmal, 1969, fiel die Ansprache aus), und zwar tatsächlich im Abstand von ziemlich genau 12 Monaten, also immer auch zur selben Jahreszeit (wodurch klimatische Einflüsse auf die Stimme der Königin dennoch nicht auszuschließen sind, da einige wenige Aufnahmen nicht in Großbritannien, sondern während Reisen innerhalb des Gebiets des Commonwealth, also in zum Teil völlig unterschiedlichen Klimata aufgezeichnet wurden), und zweitens bieten diese Daten zwar keine Laborsprache, aber durch das jährlich wiederkehrende Thema Weihnachten und Neujahr doch einen eingeschränkten und vor allem wiederkehrenden Wortschatz. Da diese Aufnahmen gekauft werden müssen, konnte zwar nicht jeder Jahrgang besorgt werden, aber doch ein Ausschnitt, der aus jedem Jahrzehnt seit der ersten Ansprache 1952 bis hin zur letzten verfügbaren aus dem Jahre 2002 mindestens drei Jahre (aus den siebziger Jahren), meistens aber mehr, abdeckt (Details siehe in 2.2.4 auf Seite 58).

Schwieriger gestaltete sich die Suche nach geeignetem Material für einen männlichen Sprecher. Zwar gibt es zahlreiche Männer, die regelmäßig in Rundfunk oder Fernsehen zu hören waren, doch stellt sich immer wieder das Problem der Vergleichbarkeit der Aufnahmesituation und des Sprechstils. Eine Ausnahme von dieser Regel bildet der seit den dreißiger Jahren des zwanzigsten Jahrhundert bis hin zu seinem Tode im Jahre 2004 für die BBC sendende Journalist Alistair Cooke, der seit Mitte der Vierziger Jahre wöchentlich in den USA eine ca. viertelstündige Sendung (*Letter from America*)produzierte und damit

bis 2004 fortfuhr. Es gibt hierbei zwei Einschränkungen zu beachten: Cooke, ursprünglich Brite und wohl auch RP-Sprecher (er war in den dreißiger Jahren sogar in der Kommission in der BBC, die die ‚korrekte' Aussprache vorschreiben sollte) lebte während des gewählten Zeitraums in den USA, und damit ist seine Sprache Einflüssen der General American genannten Varietät des Englischen ausgesetzt gewesen; rein impressionistisch läßt sich feststellen, dass diese Beeinflussungen seines Idiolekts mit der Zeit variierten, so dass zumindest eine detailierte Analyse z. B. seiner Vokalräume wohl ausgeschlossen werden müsste, wollte man soziophonetische und akzent-kontakt-bedingte Variation ausschließen; da Einzelbetrachtungen auf Phonemebene (jenseits des Schwa-Vokals aus Experiment 1.1) für diese Studie ohnehin nicht in Frage kommen, kann es aufgrund der Einzigartigkeit seiner Aufnahmen als Langzeitkorpus in Kauf genommen werden, dass solche phonetische Variation in seiner Sprache erwartet werden kann. Zweitens ist der Themenkreis seiner Sendungen zwar beschränkt auf politische, historische, gesellschaftliche und soziale Fragestellungen, dennoch ist etwas mehr Variabilität zu erwarten als in den Weihnachtsansprachen der Königin, in denen teilweise sogar die selben Wörter (wie z. B. *christmas*, *merry*, *year*, usw.) in verschiedenen Jahren vorkamen.

Ein Vorteil von Cookes Aufnahmen gegenüber denen der Königin ist es, dass nicht nur für jedes Jahr 5 bis 15 Minuten vergleichbares Sprachmaterial in bester Studioqualität zur Verfügung stehen, sondern theoretisch 10 bis 15 Minuten für (fast) jede Woche seit 1947 bis Anfang 2004. Selbstverständlich setzen die finanziellen Mittel auch hierbei Grenzen; dennoch konnten für Cooke zehneinhalb Stunden Sprachdaten, für die Königin jedoch 'nur' zweieinhalb Stunden Gesamtmaterial analysiert werden.

Methode

Berechnet wurden f0 und Formantfrequenzen in 29 Weihnachtsansprachen der Königin und in 47 Sendungen von „Letter from America" aus 30 Jahren von Cooke. Diese Aufnahmen stammten bei der Königin aus den folgenden Jahren: 1952, 1954-59, 1960, 1962-68, 1970-72, 1983, 1985, 1988, 1994-99, 2000-02. Für Cooke ist Material aus den Jahren 1947, 1951, 1953, 1960, 1962, 1965, 1970-71, 1973-74, 1980-85, 1990-94, 1996-99, und 2000-2004 verfügbar. Die durchschnittliche Dauer der Aufnahmen betrug 5,2 Minuten bei der Königin und 13,5 Minuten bei Cooke, was Gesamtdauern von 2 Stunden, 35 Minuten für die Königin bzw. 10 Stunden, 35 Minuten für Cooke ergibt. Diese enorme Datenmenge ließ es unwirtschaftlich erscheinen, die Methoden aus dem ersten Experiment, also die Etikettierung in Schwa-Segmente, anzuwenden, zumal nur für eine Teilmenge der Aufnahmen Cookes' Transliterationen (und diese nur in Form von nachträglich redigierten – und somit möglicherweise vom tatsächlich gesprochenen abweichenden – Sendungsmanuskripten) vorlagen, die es erlaubt hätten, automatische Segmentationssysteme wie beispielsweise *MAuS* (Kipp, 1999; Beringer & Schiel, 1999; Schiel, 1999, 2004) anzuwenden. Stattdessen wurden wie in Experiment 1.3 Formant- und Grundfrequenzwerte in allen stimmhaften Signalabschnitten ermittelt und diese pro Aufnahmeausschnitt (Dauer: circa eine Minute) gemittelt, und diese Werte wiederum pro Aufnahmejahr gemittelt, sodass sich für f0 und jeden Formanten jeweils ein Wert pro Jahr ergab.

Die Mittelung pro Jahr wurde vorgenommen, um mittels Linearer Regression den altersbedingten Wandel in f0, F1, F2 und F3 als Funktion des Alters darstellen zu können. Untersucht wurde, ob die Rate des Wandels linear verläuft oder sich im Laufe der Zeit verändert.

Ergebnisse

Der natürliche Logaritmus (zur Basis e, der Euler'schen Zahl) der Parameter f0, F1, F2 und F3 erwies sich als jenes Maß, welches die systematischsten Zusammenhänge zwischen den Parametern und Sprecheralter aufwies. Diese Zusammenhänge sind in Abbildung 2.11 als lineare Abhängigkeit vom Alter bei den Daten der Königin ersichtlich, und, für f0 und den ersten Formanten aufgeteilt in zwei getrennt berechnete Zeiträume, als zwei lineare, in ihrer Richtung entgegengesetzt verlaufenden Abhängigkeiten bei Cooke (F3 und F4 können auch bei ihm linear modelliert werden, siehe Abbildung 2.12). Beide Abbildungen enthalten normalisierte, also um die parametereigene Mittellage zentrierte Daten, um eventuell auftretende Ähnlichkeiten in den Verläufen besser einschätzen zu können. Bei der Königin wird F3 nicht durch das Alter erklärbar, was auch daran zu sehen ist, dass die Regressionslinie praktisch auf der Nulllinie liegt ($R^2 = 0,01; F[1,25] = 1,3$, n.s.); F2 fällt leicht mit dem Alter ($R^2 = 0,38; F[1,25] = 16,7; p < 0,01$); stärker fallen F1 ($R^2 = 0,76; F[1,25] = 82,4; p < 0,01$) und f0 ($R^2 = 0,71; F[1,25] = 63,5; p < 0,01$). Da es aussieht, als fielen f0 und F1 in der gleichen Rate, wurde untersucht, ob die Steigungen der Regressionslinien für f0 und F1 sich signifikant unterscheiden, indem die Methode aus Pedhazur (1997) zum Vergleich zweier Regressionslinien angewendet wurde. Die Regressionslinien fallen (f0 mit $-0,0046\ ln(Hz)/Jahr$ und F1 mit $-0,0054\ ln(Hz)/Jahr$) mit nicht signifikant unterschiedlicher Rate ($F[1,50] = 0,85$; n.s.).

Auch für Cooke erklärt das Alter F3 nicht ($R^2 = 0,0008; F[1,28] = 0,02$; n.s.); F2 steigt minimal mit dem Alter an ($R^2 = 0,2; F[1,28] = 6,84; p < 0,05$), fällt also nicht wie bei der Königin. Wie oben bereits erwähnt, ließen sich F1 und die Grundfrequenz dann linear modellieren, wenn man die Daten in zwei Abschnitte teilte, da vor einem gewissen Alter beide Parameter zu sinken, und nach diesem Alter beide anzusteigen schienen. Dieser Umkipppunkt wurde nicht grob durch Dateninspektion durch den Autoren abgeschätzt, sondern berechnet mithilfe einer in Kapitel 6 in Baayen (2009) beschriebenen Methode, die jenen Punkt findet, an dem die Summe der quadrierten Abweichungen zweier linearer Regressionen, die auf einen Datensatz angewendet werden, minimiert werden. Für die Grundfrequenz Cookes angewendet, ergibt sich das 48ste Jahr des Messzeitraums (also als Cooke das Alter von 87 Jahren erreicht hatte) als Umkipppunkt, für F1 ergibt sich ein Umkipppunkt, der nur ein Jahr später liegt, also als Cooke 88 Jahre alt war. Da dieser Umkipppunkt in beiden Fällen sehr ähnlich war, liegt die Vermutung nahe, dass auch bei Cooke eine Kovariation zwischen f0 und F1 zu erwarten ist, auch wenn die Abbildung 2.12 darauf schließen lässt, dass F1 vor dem Umkipppunkt stärker als f0 fällt. Um mittels Pedhazurs Methode (Pedhazur, 1997) die Raten des Abstiegs/Anstiegs beider Parameter f0 und F1 miteinander vergleichen zu können, mussten die Regressionen bis zu einem einzigen Umkipppunkt ermittelt werden, und wegen der Ähnlichkeit beider vorher

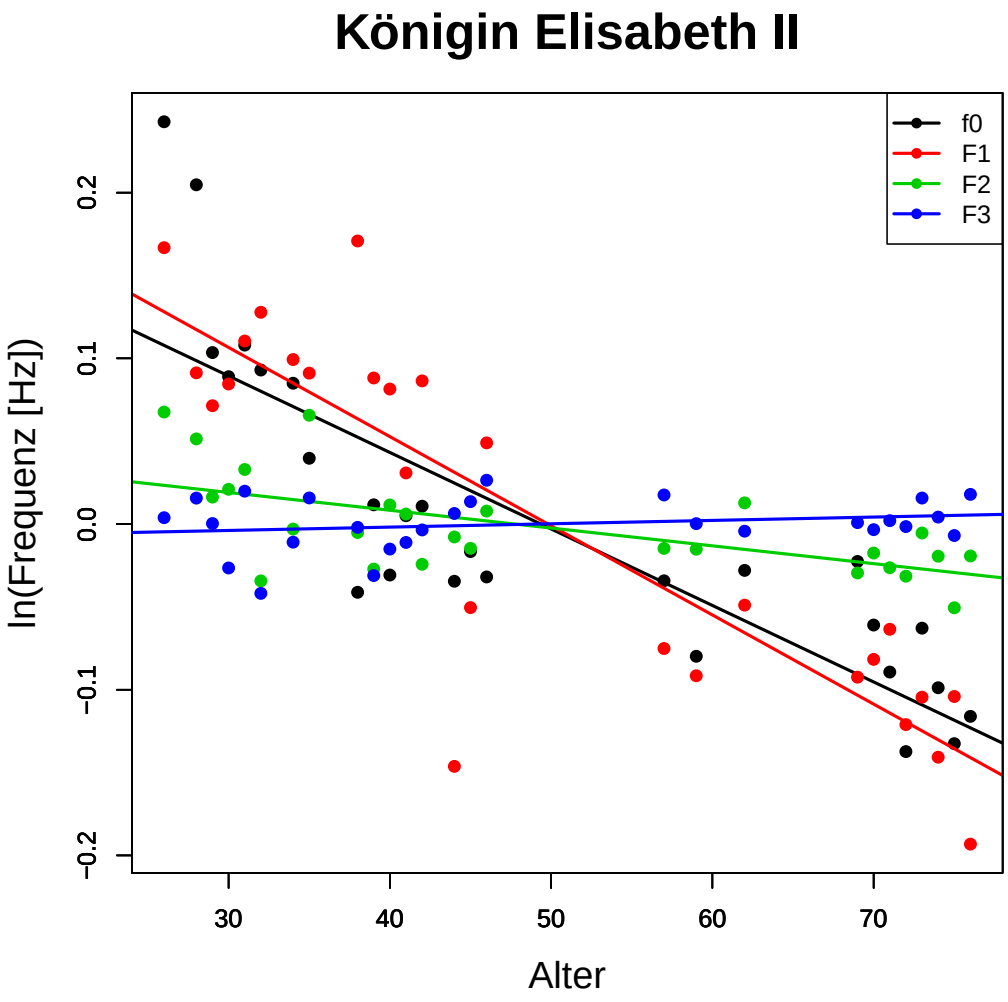

Abbildung 2.12: *Die um den eigenen Mittelwert reduzierten, also normalisierten f0- (schwarz), F1- (rot), F2- (grün), und F3- (blau) -Werte, dargestellt als natürlicher Logarithmus zur Basis e ihrer Hertz-Werte, bei Königin Elisabeth II, als Funktion ihres Alters. Die Linien stellen die linearen Regressionen dar. Details zu den Regressionen: siehe Text in 2.2.4*

ermittelten Umkipppunkte für f0 (87) und F1 (88) erschien dies als vertretbar, Cookes 87stes Lebensjahr als gemeinsamen Umkipppunkt auszuwählen. Während tatsächlich die Steigungen nach dem Umkipppunkt für f0 ($R^2 = 0,88; F[1,7] = 53,2; p < 0,05$; $+0,0298$ $ln(Hz)/Jahr$) und F1 ($R^2 = 0,73; F[1,7] = 19,15; p < 0,05$, $+0,0269$ $ln(Hz)/Jahr$) nicht signifikant unterschiedlich sind ($F[1,14] = 0,15$; n.s.), fällt f0 vor dem Umkipppunkt ($R^2 = 0,69; F[1,19] = 42,2; p < 0,005$; $-0,0042$ $ln(Hz)/Jahr$) offenbar deutlich flacher ($F[1,38] = 9,45; p < 0,01$) als F1 ($R^2 = 0,76; F[1,19] = 61,36; p < 0,01$; $-0,0079$ $ln(Hz)/Jahr$).

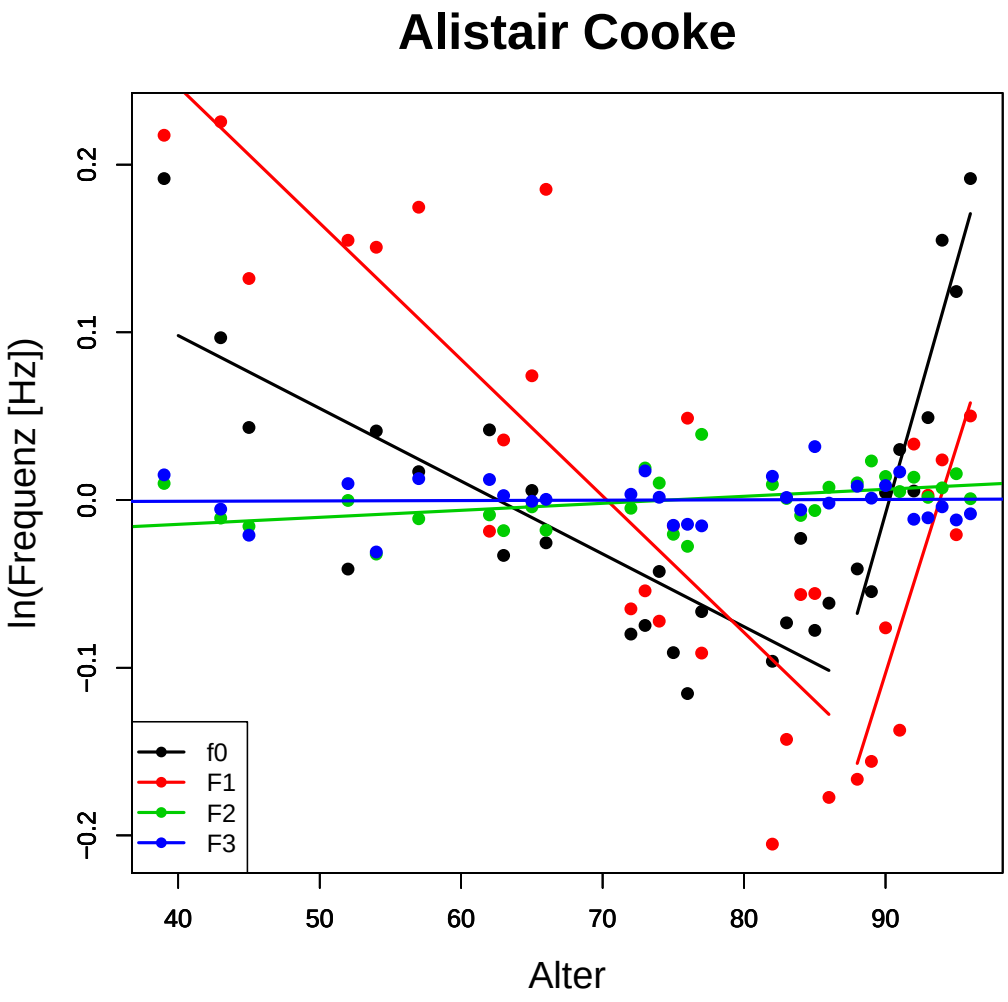

Abbildung 2.13: *Die um den eigenen Mittelwert reduzierten, also normalisierten f0- (schwarz), F1- (rot), F2- (grün), und F3- (blau) -Werte, dargestellt als natürlicher Logarithmus zur Basis e ihrer Hertz-Werte, bei Alistair Cooke, als Funktion seines Alters. Die Linien stellen die lineare Regression dar. Die Grundfrequenz und der erste Formant konnten nur abschnittsweise linear modelliert werden. Weitere Details zu den Regressionen: siehe Text in 2.2.4*

Zusammenfassend lässt sich also feststellen, dass F3 weder bei Cooke noch bei der Königin durch das Alter erklärbar sind, F2 bei der Königin leicht, aber eindeutig fällt, bei Cooke aber leicht (und weniger eindeutig) steigt; für f0 und F1 lässt sich feststellen, dass sie mit dem Alter bei beiden Sprechern offenbar kovariieren; während bei der Königin Elisabeth II Grundfrequenz als auch der erste Formant in einer sehr ähnlichen Rate linear absinken, fallen beide Parameter zunächst bei Cooke (allerdings in unterschiedlichen Raten), um nach einem bestimmten Alter (bei Cooke 87/88 Jahre) wieder anzusteigen, und zwar in

der gleichen Rate.

Um zu verdeutlichen, dass durch diese Kovariation der Abstand zwischen *F1* und *f0* bei beiden Sprechern auf einer logarithmischen Skala über die Jahrzehnte realtiv invariant bleibt, präsentieren wir eine alternative Abbildung der gleichen Daten (nur für die genannten zwei Parameter), entnommen aus Reubold et al. (2010).

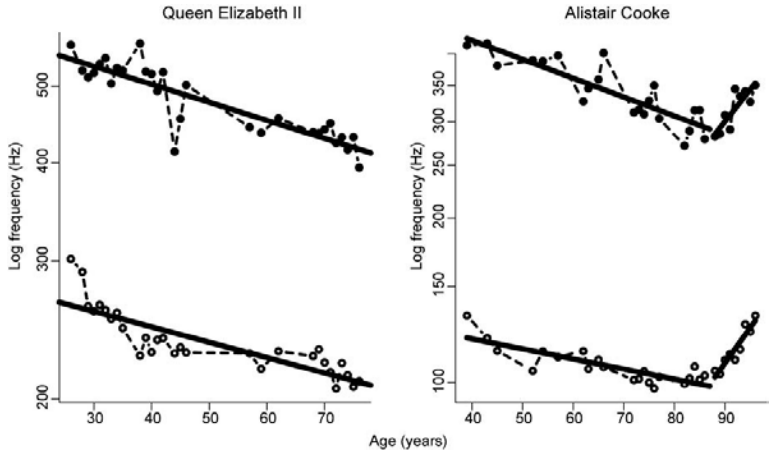

Fig. 3. Mean ln F_0 (unfilled circles) and mean ln F_1 (filled circles) in voiced frames for Queen Elizabeth II (left) and Alistair Cooke (right) as a function of chronological age with superimposed regression lines through the scatter.

Abbildung 2.14: *f0 und F1 über die Jahrzehnte bei der Königin und Cooke. Abbildung zitiert nach Reubold et al. (2010, Seite 642), mit originaler Abbildungsunterschrift.*

Diskussion der Ergebnisse

Die Analyse zu mehr als nur zwei Zeitpunkten brachte für die Daten der Königin nur vergleichsweise geringen Erkenntnisgewinn, wenn man die Daten isoliert betrachten würde. Zwei Dinge, die an den zweistufigen, dreißig Jahre auseinanderliegenden Aufnahmen bereits abzusehen waren, sind, dass f0 mit dem Alter sinkt, und dass F1 mit dem Alter sinkt; neu und nicht aus der zweistufigen Analyse abzusehen war, dass F2 leicht mit dem Alter sinkt. Dieser Befund war aber bereits in Harrington (2006) bei einer zweistufigen und in Harrington et al. (2007a) bei einer mehrstufigen (mit weniger Stufen (nämlich: eine Stufe pro Jahrzehnt) als in der vorliegenden) Analyse für die Königin konstatiert worden. F3 bleibt - selbstverständlich mit einer gewissen Variation - über die Jahrzehnte gleich. Sehr viel mehr lässt sich - wenn man nur die Daten der Königin betrachtet - nicht ablesen. Cookes Daten öffnen jedoch die Augen für einen Zusammenhang, der bislang so nicht unbedingt erkennbar war, und der auch den Betrachter dazu zwingt, die Daten der Königin nochmals genauer zu untersuchen. Bei Cooke gibt es wiederum keine Alterseffekte für F3, aber leicht mit dem Alter steigende F2-Werte. Dies lässt Hypothesen, die bei den spektralen Änderungen über das Alter von einen globalen Längung des Vokaltraktes ausgehen (Linville, 2001), unwahrscheinlicher werden. Entscheidender ist jedoch, dass nicht nur das eigentliche, ursprüngliche Ziel dieses Experiments - nicht-lineare Änderungen der Grundfrequenz -

die zunächst abfällt und nach einem gewissen Alter wieder ansteigt, aufzudecken - erreicht wurde, sondern auch ein erstaunlicher Befund zu Tage trat, der einen wie auch immer gearteten Zusammenhang von Grundfrequenz und erstem Formanten nahelegt: Nicht nur die Grundfrequenz steigt nach einem gewissen Alter (wie von der Literatur (Baken, 2005; Linville, 1996, 2001) vorhergesagt und als Hypothese in der Diskussion der Ergebnisse zu Experiment 1.3 vorgeschlagen), sondern in etwa zum gleichen Zeitpunkt ändert der bislang ebenfalls fallende erste Formant auch seine Richtung und steigt im weiteren Verlauf. Das Wiederansteigen des ersten Formanten im höheren Alter (zumindest bei Männern) ist nach Wissen des Autoren noch nicht beschrieben worden, wenn man von den berichteten *F1*-Anstiegen in Torre III und Barlow (2009) bei bestimmten Vokalen absieht (im gleichen Paper war auch von einer für ältere Männer steigende Grundfrequenz die Rede). Auch die F1-Daten der zwei Männer aus Experiment 1.3, deren f0 altersabhängig stieg, scheinen einen Anstieg von F1 im höheren Alter nicht zu bestätigen, da sie bei beiden Versuchspersonen eindeutig fallen. Da die Wende zu einem Anstieg von f0 und F1 bei Cooke zu etwa der gleichen Zeit stattzufinden schien, war es wichtig, diesen Umkipppunkt genauer zu bestimmen, als eine grobe Schätzung dies zulässt. Um die Daten vor und nach der Wende in einen linearen Zusammenhang mit der Variable Alter zu bringen, wurde das Maß der (natürlichen)Logarithmisierung der f0- und F1-Werte (in Hertz) gewählt, wobei festgestellt werden konnte, dass nach dem – nun arithmetisch bestimmten – Wendepunkt in Cooke f0 und F1 sogar die Rate des Wandels für beide Parameter gleich war. Vor dem Wendepunkt war zwar die Rate (die ja immer auch von der Transformation der Werte abhängig ist) nicht identisch, aber der Verlauf doch zumindest in die gleiche Richtung weisend. Auf die Daten der Königin angewendet, ergab sich durch Logarithmisierung der Hertz-Werte nun auch nicht allein ein lineares Absinken von f0 und F1 mit dem Alter, sondern auch die Erkenntnis, dass dieser Alterseinfluss auf beide Parameter in etwa gleich ist. Man kann also durchaus eine Kovariation der Parameter f0 und F1 hypothetisieren. So ergibt sich, wenn man diese Ergebnisse betrachtet, eine neue Betrachtungsweise: eventuell ändern sich die Formanten nicht als Konsequenz eines sich ändernden Ansatzrohrs (und damit, akustisch betrachtet, des Filters), sondern als Konsequenz eines sich veränderten Quellsignals - oder die Glottalschwingung ändert sich aufgrund von Änderungen im Ansatzrohr. Es sind aus der Literatur mehrere physiologische, akustische und perzeptive Gründe entnehmbar, warum es denkbar ist, dass Quelle und Filter nicht so unabhängig voneinander sind, wie man in einfachen frühen Modellen der akustischen Modellierung des Sprachsignals annehmen durfte. Bevor jedoch diese Gründe und Zusammenhänge im einzelnen ausführlich besprochen und - sofern möglich - zu hypothetischen Annahmen formuliert und diese getestet werden sollten, muss im folgenden Experiment eine Möglichkeit ausgeschlossen werden, die die hier beschriebenen Zusammenhänge auch erklären könnte: die Möglichkeit, dass es sich ausschließlich um Artefakte der akustischen Messungen handelt.

2.2.5 Ist der Zusammenhang zwischen f0 und F1 ein Artefakt der akustischen Berechnungen?

Dieses Kapitel ist mehr oder weniger unverändert aus (Reubold et al., 2010) entnommen; um dem Leser den Übergang zwischen den Teilexperimenten leichter und vor allem plausibler zu machen, ist es jedoch notwendig, eine mögliche technische Fehlerquelle auszuschließen, was in diesem Kapitel getan wird.

Harrington (2006) wies bereits darauf hin, dass ein gewisser Einfluss der Grundfrequenz auf die Messergebnisse durch die LinearPredictiveCoding-(LPC)-Methode nicht gänzlich ausgeschlossen werden kann, da die LPC-Methode zwar genau darauf beruht, Quelle und Filter zu trennen, andererseits das Quellsignal aber über die Harmonischen Einfluss darauf hat, wo das Frequenzband „Formant“ seinen Intensitätsschwerpunkt hat. Dies gilt insbesondere für den ersten Formanten, bei dessen gewöhnlicher Lage die Harmonischen, die mit zunehmender Frequenz gewöhnlich an Amplitude abnehmen, noch relativ große Intensität aufweisen; zusätzlich ist zu beachten, dass wegen der relativ geringe Bandbreite des ersten Formanten dieser am anfälligsten ist für Einflüsse, die durch die Harmonischen der Grundfrequenz entstehen („because F1 bandwidth is usually small, the error range will be quite large“ (Vallabha & Tuller, 2002). Um zu testen, ob solche Einflüsse der Grundfrequenz auf Messergebnisse der LPC-Analyse einen Einfluss haben, wurden Sprachsignale in Quell- und Filtersignal getrennt, nur der Quellanteil manipuliert und durch Resynthese mit dem ursprünglichen Filtersignal wieder zu einem Sprachsignal zusammengefügt. Die anschließend gemessenen F1-Werte sollten sich in den verschiedenen Resynthesen idealerweise nicht signifikant voneinander unterscheiden.

Methode

Als Material für die Resynthesen wurden Schwa-Vokale aus jenem Lebensjahr der Königin gewählt, das für das Maß der Grundfrequenz die Mitte des Umfangs während des Beobachtungszeitraums bildete, nämlich Daten aus dem Jahr 1960 mit einer mittleren Grundfrequenz von 260 Hz. Diese Mitte der Verteilung wurde gewählt, um die Grundfrequenz dergestalt manipulieren zu können, dass ähnlich Werte wie zu Beginn und Ende der Beobachtung (1952: ca. $310Hz$; 2002: ca. $210Hz$) resynthetisiert werden konnten, ohne in unnatürliche Bereiche vorzustoßen, und so die gesamte Verteilung der in den Daten der Königin vorkommenden f0-Werte abzudecken. Als weiterer Grund für die Auswahl von Daten aus dem Jahre 1960 muss genannt werden, dass für dieses Jahr die meisten Schwas segmentiert vorlagen ($N = 200$). Für die Resynthesen wurde die Implementierung der PSOLA-Methode (Moulines & Charpentier, 1990) in Praat (Boersma & Weenink, 2010) gewählt. Diese trennt Quelle und Filter und setzt beide Signalanteile zur Resynthese wieder zusammen, wobei die Möglichkeit besteht, das Quellsignal, insbesondere die Grundfrequenz, getrennt zu manipulieren. Jeder der 200 Schwas wurde mit den Original-Werten, sowie mit $\pm 10\%$ und $\pm 20\%$ der originalen f0-Werte resynthetisiert, so dass für f0-Mittelwerte der Bereich von ca 200 bis ca 310 Hz abgedeckt werden konnte. Anschließend wurden in den Resynthese mit den gleichen Einstellungen wie in Experiment 1.1 f0- und F1-Werte mittels f0ana und

forest gemessen. Mittels statistischer Test (Details siehe 2.2.5) wurde ermittelt, ob sich F1-Werte in den grundfrequenzmanipulierten Resynthesen unterschieden.

Ergebnisse

Abbildung 2.15 zeigt das Ergebnis der Messungen für f0 und F1 in den f0-manipulierten Resynthesen. Zwischen der +20%-Bedingung und der −10%-Bedingung fällt F1, ebenso wie f0, wenn auch sicherlich nicht in der gleichen Rate; von der −10%- zur −20%-Bedingung steigt F1 jedoch wieder an, obschon f0 weiterhin fällt. Eine Varianzanalyse mit Messwiederholung mit F1 als abhängiger Variable, einem geordneten Faktor mit den Stufen von −20% bis +20% als unabhängige Variable *Resynthesebedingung* sowie mit Schwa als Fehlervariable (da jeder der 200 Schwas unter fünf Bedingungen gemessen wurde) ergab einen signifikanten Effekt der *Resynthesebedingung* ($F[4, 796] = 22, 6; p < 0, 01$). Ein mit Schwa als Zufallsvariable durchgeführte Analyse mittels eines Linearen Gemischten Modells ergab für F1 sowohl einen linearen ($t = 6.2$) als auch einen kubischen ($t = 6.5$) Trend. Dies lässt sich in Übereinstimmung bringen mit der vorher durchgeführten Inspektion der Abbildung der Daten in Abbildung 2.15; insgesamt ist ein leicht fallender Trend zu beobachten, jedoch weist der Verlauf auch kubische Eigenschaften auf, da gegen Ende F1 wieder ansteigt.

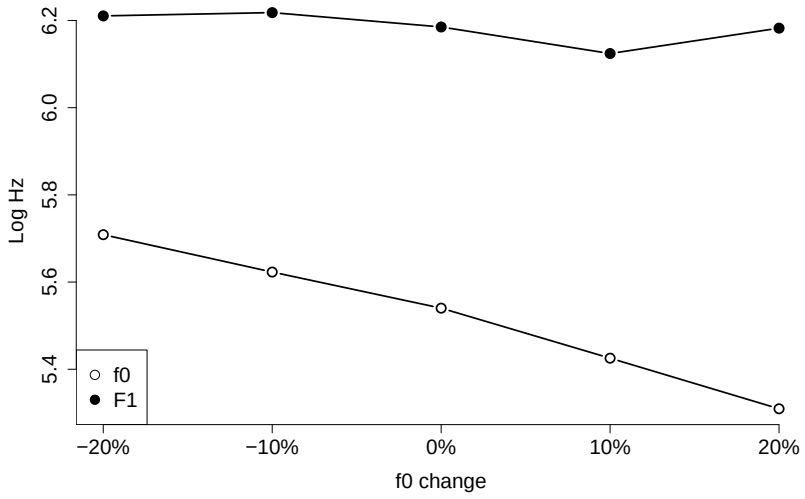

Abbildung 2.15: *Gemittelte f0- (leere Kreise) und F1- (gefüllte Kreise) Werte in den Schwas aus der Weihnachtsansprache 1960 der Königin Elisabeth II. Die originalen Werte sind bei '0%'; die Labels '±10%' und '±20%' verweisen auf jene Werte, die in Resynthesen, bei denen entsprechend der Labels f0 manipuliert worden war, gemessen wurden. Zitiert aus Reubold et al. (2010)*

Diskussion der Ergebnisse

Es ist zu beachten, dass das hier ermittelte Muster für F1-Messwert-Änderungen durch Manipulation der f0-Werte völlig anders ist als das in Experiment 2.1 anhand natürlicher Sprache ermittelten Ergebnissen, so dass sowohl Reubold et al. (2010) als auch diese Arbeit davon ausgeht/ ausgehen wird, dass diese Artefakt-Effekte zu vernachlässigen sind, da sie von anderer Gestalt und wesentlich geringerer Quantität sind als die Zusammenhänge, die in Experiment 2.1 offenbar wurden. So fällt z. B. in Experiment 2.1 F1 bei Cooke vor dem Wendepunkt um einiges stärker als f0; von einem solchen Muster sind die hier vorliegenden Daten weit entfernt.

2.2.6 Wie verändern sich die akustischen Korrelate des Open Quotient und der Behauchung?

Zusammenhang zwischen F1 und laryngealen Maßen

Bevor wir in der allgemeinen Diskussion mehrere Hypothesen aufstellen werden, wie es zu einer Kovariation von Grundfrequenz und erstem Formanten kommen könnte, wollen wir vorgreifen und akustische Korrelate einer der möglichen Quellen für diesen Zusammenhang bestimmen. Diese mögliche Quelle besteht darin, dass zwar akustische Quelle und akustischer Filter in einer Approximation zur Darstellung und Berechnung eines Sprachsignals als getrennt voneinander betrachtet werden können, aber dennoch bekannt ist, dass zwischen Quelle und Filter Interaktionen bestehen. So kann man sagen, dass das Quellsignal durchaus einen Einfluss auf den Filter und hierbei vor allem auf F1 hat (da während der Öffnungsphase der Glottis ein weiterer Resonator angekoppelt ist) und umgekehrt der Filter das Verhalten der Glottisschwingung beeinflussen kann (Childers & Wong, 1994; Klatt & Klatt, 1990). Diese Interaktionen werden in der allgemeinen Diskussion eingehender besprochen; wichtig für das folgende Experiment ist nur, dass es Hinweise aus der Literatur gibt (Barney, de Stefano & Henrich, 2007), dass F1 während der Öffnungsphase wesentlich höhere Werte (13%) als während der geschlossenen Phase annehmen kann (neben Änderungen seiner Amplitude und Bandweite), so dass von höheren F1-Werten für Sprachsignale ausgegangen werden muss, wenn der gleiche Filter von einer Quelle mit einen höheren Open Quotient (Verhältnis von offenen Phasen zu geschlossenen Phasen der Glottis, gemessen als prozentualer Anteil der offenen Phase innerhalb der Gesamt-Periode) angeregt wird. Eine weitere Überlegung ist, dass, wenn eine (teilweise) geöffnete Glottis einen höheren ersten Formanten verursachen kann, dies auch für behauchte Stimmgebung gelten sollte, da in beiden Fällen die über die Schwingungsperiode gemittelte Querschnittsfläche der glottalen Öffnung erhöht sein sollte. Umgekehrt betrachtet, und in Hinblick auf die vorliegenden Formantdaten entscheidender, sollte eine gepresstere Phonation zu einem Absinken des ersten Formanten führen; man beachte jedoch, dass hierzu auch gegenteilige Evidenz vorliegt, denn Ladefoged, Maddieson und Jackson (1988) finden einen höheren ersten Formanten für gepresst phonierte Vokale des Mazatec (wo modale und gepresste Phonation distinktive Merkmale sind und deshalb möglicherweise anders, nämlich aktiv produziert werden,

während die hier zu untersuchende Gepresstheit als passives Altersphänomen angenommene wird), was sie auf ein Anheben des Kehlkopfes zurückführen. Gehen wir jedoch nur von der durchschnittlichen Querschnittsfläche der Glottis aus, so sollte unsere Grundannahme dennoch haltbar sein.

Abbildung 2.16: *Glottis während gepresster (links) und behauchter (rechts) Phonation. Zitiert aus Ladefoged (2005). Deutlich zu erkennen ist der Unterschied in der Querschnittsfläche der glottalen Öffnung; die Ankopplung der subglottalen Räume, gebildet durch die Trachea, ist somit bei behauchter Phonation wesentlich deutlicher ausgeprägt.*

Sowohl für den Open Quotient als auch für (die perzeptiv wahrgenommene) Behauchung gibt es akustische Korrelate (so z. B. die relative Amplitude der ersten Harmonischen (Hillenbrand, Cleveland & Erickson, 1994) für die Behauchung); wir wollen hier zwei davon nutzen und anhand der vorliegenden Daten bestimmen. Der Open Quotient wird in mehreren Studien mit dem Maß $H1*-H2*$ [in dB] assoziiert, also der Relation zwischen den Amplituden der ersten und der zweiten Harmonischen, wobei beide wegen des Einflusses des ersten Formanten korrigiert werden müssen, um reliabel und vor allem um über Vokale und unterschiedliche Sprecher hinweg vergleichbar zu sein (die Korrektur wird durch $*$ gekennzeichnet) (Hanson, 1997; Hanson & Chuang, 1999; Iseli, Shue & Alwan, 2007). Laut

Hanson (1995) und Jiang, Tao und Cai (2002) ist hingegen das Maß $H1*-A3*$ [in dB][10] als Maß für den sogenannten *spectral tilt* am besten mit (der wahrgenommenen) Behauchtheit korreliert.

Hier wird VoiceSauce (Shue, 2011; Shue, Chen & Alwan, 2010) benutzt werden, um beide Maße zu bestimmen. Mit verlässlichen Werten ist nur in stimmhaften Signalen mit ausgeprägten Formanten zu rechnen, also in Vokalen (Shue et al., 2010). Für beide Maße ($H1*-H2*$ und $H1*-A3*$) gilt, dass sie um so höher sind, je größer der Open Quotient bzw. die Behauchtheit ist (Fischer-Jørgensen, 1967) (dort allerdings benutzt für die linguistisch relevante, nämlich distinktive Behauchung). Ausgehend davon, dass bei allen 5 Sprechern aus Experiment 2.1. (deren Daten hier benutzt werden, da für sie Segmentationen vorliegen) F1 sank, müsste man also hypothestisieren, dass zumindest entweder $H1*-H2*$ oder $H1*-A3*$ für diese Sprecher sinken sollte, wenn die durchschnittliche Querschnittsfläche der Öffnung an der Glottis das Absinken von F1 entscheidend (mit-)beeinflusst haben sollte.

Methode

VoiceSauce wurde benutzt, um bei 5 Sprechern in 1064 Schwa-Vokalen (denselben wie in Experiment 2.2.1) $H1*-H2*$ und $H1*-A3*$ zu messen. Die Einstellungen für die Formant-Analyse, die mit den gleichen Algorithmen wie in *forest* durchgeführt wurden, waren die gleichen wie in Experiment 2.2.1. (Fensterlänge von 30 ms und einer Schrittbreite von 5 ms); f0, dessen Messung nötig ist für die Bestimmung der Harmonischen, wurde allerdings mit dem Grundfrequenzextraktionsalgorithmus aus *STRAIGHT* (Kawahara, Masuda-Katsuse & de Cheveigné, 1999), dessen Fensterverschiebung auf eine Millisekunde festgelegt ist, benutzt, da er als besser für diese Aufgabe evaluiert wurde (Shue, 2011). Im Unterschied zu Experiment 2.1. ergeben sich die jeweils 1064 Messwerte (=ein Wert pro Schwa) nicht daraus, dass nur ein Wert (am zeitlichen Mittelpunkt) extrahiert wurde, sondern dass stattdessen alle Werte eines Schwas gemittelt wurden.

Ergebnisse

Die Abbildungen 2.17 und 2.18 zeigen die Parameter $H1*-H2*$ und $H1*-A3*$ pro Sprecher sowie Interaktions-Plots. Für Cooke, Lockwood und die Königin steigt $H1*-H2*$, während für Thatcher der Wert mehr oder weniger unverändert bleibt und für Plomley sinkt. Ein Lineares Gemischtes Modell mit $H1*-H2*$ als abhängige Variable und *Alter* als Zwischensubjektfaktor sowie Ausklammerung der Sprecher ergab einen hochsignifikanten Alterseffekt ($F(1, 60) = 32, 7; p < 0, 01$). Auch für die abhängige Variable $H1*-A3*$ modelliert ergibt sich ein signifikanter Alterseffekt ($F(1, 60) = 8, 3; p < 0, 01$); wie Abbildung 2.18 jedoch zeigt, ist auch dieser nicht konsistent über alle Sprecher (bei Thatcher sinkt der Wert, und er bleibt relativ unverändert bei Plomley). Hauptsächlich jedoch muss

[10] Das Maß bestimmt die Relation der Amplituden der ersten Harmonischen, korrigiert für den Einfluss des ersten Formanten, und der Amplitudenspitze des dritten Formanten, wiederum korrigiert, aber für Einflüsse des ersten und zweiten Formanten; zu Details der Korrekturen siehe Iseli et al. (2007).

festgehalten werden, dass die Werte (mit der Ausnahme Plomley und Thatchers) in die Gegenrichtung zur hypothetisierten Richtung weisen.

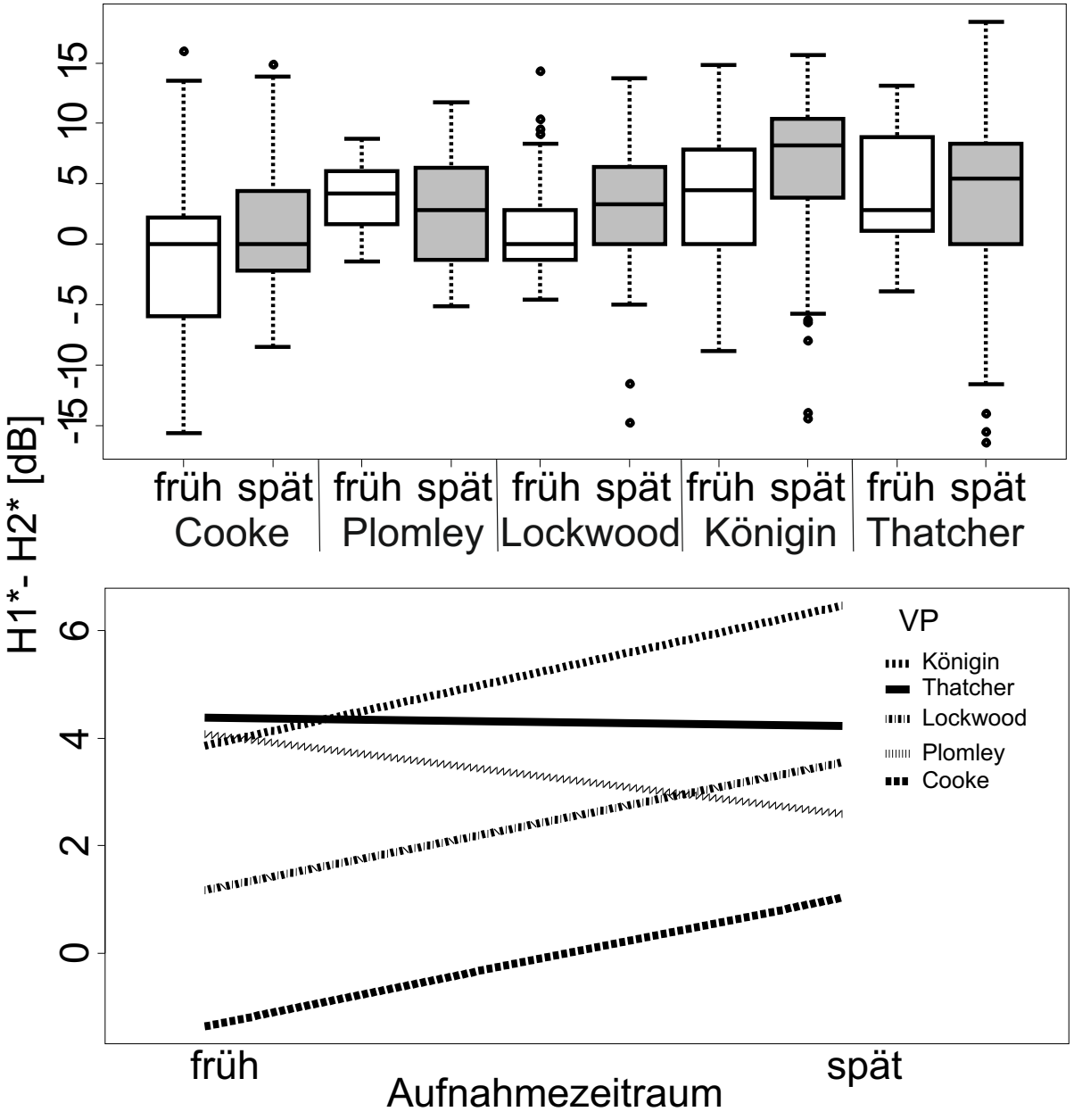

Abbildung 2.17: *Verteilung und Interaktion von $H1*-H2*$, dem akustischen Korrelat des Open Quotients, pro Sprecher für zwei Zeiträume, die circa 30 Jahre auseinander liegen.*

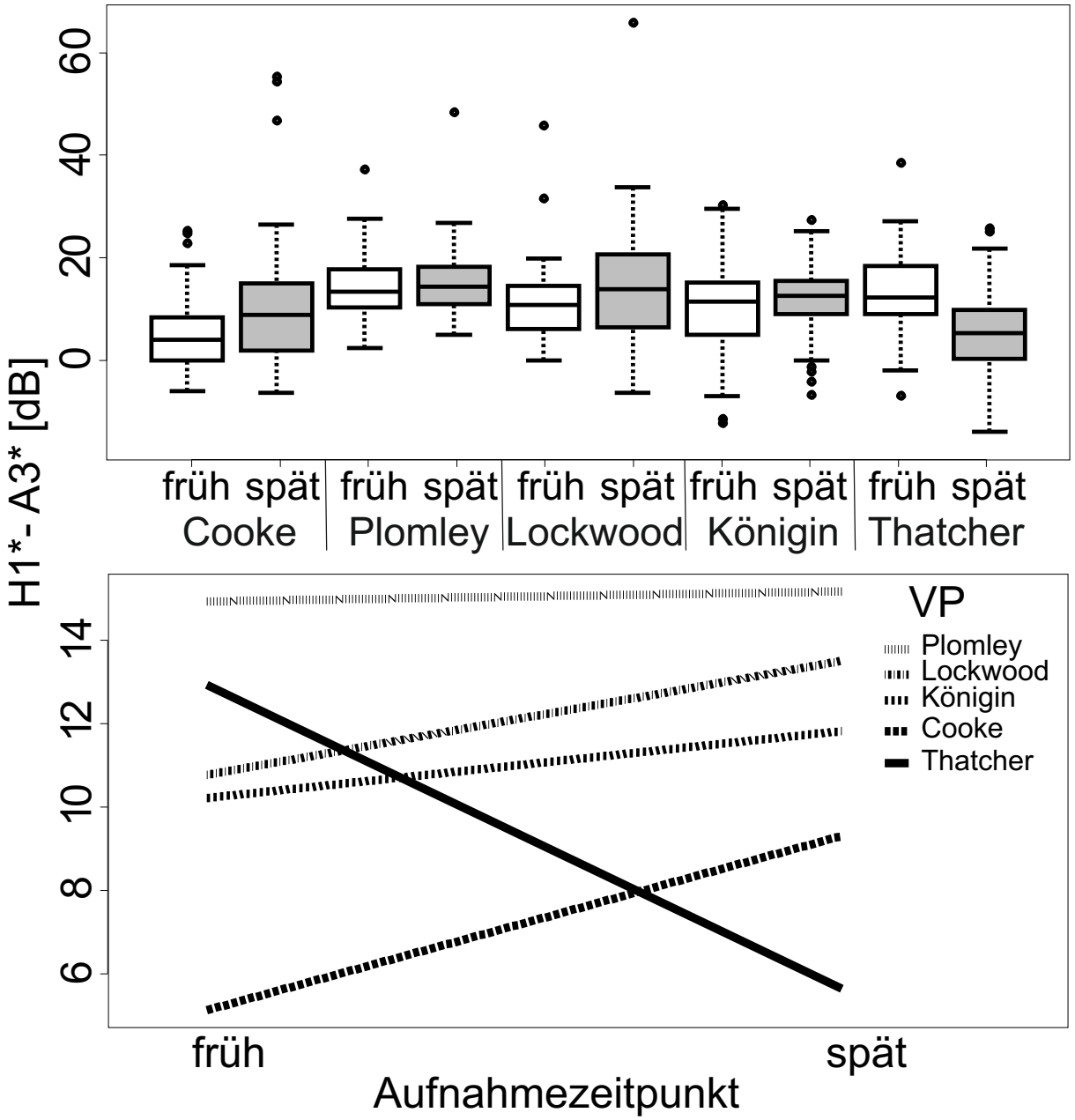

Abbildung 2.18: *Verteilung und Interaktion von $H1*-A3*$, dem akustischen Korrelat für Behauchtheit, pro Sprecher für zwei Zeiträume, die circa 30 Jahre auseinander liegen.*

Diskussion der Ergebnisse

Dieses Experiment diente nicht dazu, altersbedingte Variation in der Funktion der Stimmlippenschwingung zu untersuchen. Dazu ist umfangreiche Literatur vorhanden, wenn auch mit sich widersprechenden Ergebnissen; einen guten Überblick über Ergebnisse dieser Forschungsrichtung bietet Kapitel 7 in Linville (2001). Stattdessen sollte dieses Experiment

die Plausibilität der These testen, dass ein Abfallen des ersten Formanten in fünf Sprechern zu zwei Zeiträumen, die circa dreißig Jahre auseinanderliegen, ein Effekt der Verringerung der durchschnittlichen, über die Zeit gemittelten Querschnittsfläche der glottalen Öffnung, wie sie für gepresste Phonation erwartet werden würde, sein könnte. Es wurden keine konsistenten Ergebnisse erzielt - in beiden Experimenten gab es jeweils nur einen Sprecher/ eine Sprecherin, dessen/ deren akustisches Open Quotient- bzw. Behauchtheitskorrelat in die Richtung wiesen, die von der Hypothese vorausgesetzt wird. Stattdessen scheint die Mehrheit der Sprecher in zunehmendem Alter eher mit im Durchschnitt weiter geöffneter Glottis zu phonieren. Dies sollte, der Hypothese gemäß, eher zu einem Anstieg des ersten Formanten führen. Stattdessen sinkt der erste Formant für die gewählten Ausschnitte konsistent bei allen fünf Sprechern (vielleicht mit der Ausnahme Lockwoods), so dass es unwahrscheinlich erscheint, dass die hier ermittelten Werte überhaupt einen Zusammenhang zur altersbedingten F1-Variation aufweisen. Hätten sich Hinweise ergeben, dass F1 durch die gemittelte Querschnittsfläche der glottalen Öffnung beeinflusst worden wäre, wäre die Frage aufgetaucht, inwiefern es einen Zusammenhang zwischen der glottalen Öffnung und der Grundfrequenz der Schwingung geben könnte. Auf diese Weise wäre eine Kovariation von f0 und F1 eventuell mittelbar - also über einen Zwischenschritt - erklärbar gewesen. Da bereits der erste Einfluss ausgeschlossen werden kann, erübrigt sich an dieser Stelle eine weitere Untersuchung eines Zusammenhangs zwischen Phonationstyp und f0, da diese Fragestellung für diese Studie nun irrelevant erscheint.

2.3 Allgemeine Zusammenfassung und Diskussion

Zusammenfassend lassen sich für dieses Kapitel als die vier wichtigsten Befunde feststellen:

- *f0* variiert systematisch mit dem Alter
- *F1* variiert systematisch mit dem Alter
- *F2* und *F3* variieren nicht systematisch mit dem Alter; bei einer Sprecherin ist allerding auch für *F2* ein leichter Abfall zu beobachten
- offenbar korrelieren die Maße *f0* und *F1* positiv; also, wenn *f0* sinkt, sinkt *F1*, wenn *f0* steigt, steigt auch *F1*

Zum ersten Punkt ist zu sagen, dass im ersten Experiment, 2.2.1, bei allen Versuchspersonen *f0* sank, während im zweiten longitudinalen Experiment, in dem die Sprecherbasis um drei Männer, deren Aufnahmen ebenfalls um circa 20 bis 30 Jahre auseinanderlagen, erweitert wurde, das Bild etwas differenzierter zu betrachten ist: dort zeigen zwei Männer mit zunehmendem Alter steigende *f0*-Werte. Festzuhalten ist hierbei jedoch, dass diese zwei Männer zu einem relativ späten Alter aufgenommen wurden, und sie somit etwas älter sind als ihre Geschlechtsgenossen in dieser Studie; es ist vollkommen konsistent mit den Befunden aus der Literatur, dass die *f0* des Mannes mit dem Alter fällt, um ab einem bestimmten Zeitpunkt wieder anzusteigen (Hollien und Shipp (1972); Brown et al. (1991); Linville (1996, 2001); Baken (2005), vgl. Abbildung 2.3 auf Seite 37). Es ist also gut möglich, dass die (auch von äußeren Gegebenheiten geprägte) Auswahl der Aufnahmen, die zwar immer zwei Aufnahmen pro Sprecher mit einem Abstand von 20 bis 30 Jahren, aber eben leider zu etwas unterschiedlichen Lebensaltern enthielt, das gegebene Ergebnis beeinflusst haben könnte; eventuell waren die Sprecher, bei denen *f0* mit dem Alter stieg, im Gegensatz zu den anderen Männern schlicht schon bereits jenseits des Wendepunktes, nach dem die Grundfrequenz wieder alterbedingt ansteigt. In Teilexperiment 2.2.4 zeigten wir selbst, dass die Daten des Sprechers Cooke, der in den ersten beiden longitudinalen Experimenten jeweils fallende Werte aufweist, später ebenfalls noch einen als Umkipppunkt zu bezeichnenden Zeitpunkt zeigt, nach dem seine Grundfrequenz wieder anstieg. Allerdings fällt ein Unterschied zur Literatur auf: der Umkipppunkt bei Cooke liegt später, als in der Literatur zu finden ist, und auch bei den anderen männlichen Sprechern, deren Grundfrequenz alterbedingt sinkt, ist der in der Literatur angegebene Zeitraum für die Wende vom Abfall der *f0* zu ihrem Anstieg eigentlich schon überschritten, vielleicht mit der Ausnahme Attenboroughs.

Linville beschreibt auf den Seite 172/173 ihres Buches (Linville, 2001), dass die mittleren Grundfrequenzen bei männlichen Populationen zunächst vom Alter von 20 bis 40/50 Jahren absinken, während sie nach dem vierten Lebensjahrzehnt wieder ansteigen. Da es sich bei diesen Daten um eine Kompilation mehrerer Studien handelt, ist es nicht eindeutig, wo genau der Umkipppunkt zu bestimmen ist und wie stark die interindividuelle Variation zwischen männlichen Sprechern einzuschätzen ist. Baken (2005) beschreibt sogar einen

früheren Umkipppunkt vor dem 40sten Lebensjahr. Auch bei Brown et al. (1991) ist eine Kompilation von Daten zu finden, wo für Männer die tiefsten Werte beim Alter von 40 zu finden sind; allerdings lassen sich die Daten für Männer in Brown et al. (1991) nicht in zwei linear modellierbare Abschnitte einteilen; ein deutlicher Anstieg findet dort dann auch erst jenseits der 70 statt. Hollien und Shipp (1972), die einzige Quelle wo die Datensammlung nicht auch einer Kompilation fremder Daten, sondern aus der Auswertung der Sprache von 175 Männern im Alter zwischen 20 und 89 besteht, vermeldet ebenso einen Abstieg bis zum 40sten Lebensjahr, und danach einen Anstieg.

Es ist also gut möglich, dass erstens das Alter, zu dem man den Umkipppunkt erreicht, stark vom Sprecher und seinen Gewohnheiten abhängig ist. Da wir es im vorliegenden Fall zumeist mit mehr oder weniger professionellen Stimmbenutzern zu tun haben, ist gut vorstellbar, dass sie zumindest ansatzweise ein professionelles Stimmtraining absolviert haben, dass die Auswirkungen altersbedingter Veränderungen auf ein späteres Alter verschob. Für (die sicher stimmtechnisch wesentlich besser ausgebildeten) Sänger in der Studie von Brown et al. (1991) (*nicht* jene Sprecher , deren Daten in der Abbildung 2.3 auf Seite 37 zu sehen sind) konnten jedenfalls keine signifikanten altersbedingten Unterschiede in der Grundfrequenz gefunden werden - was dafür spricht, dass Stimmtraining altersbedingte Änderungen wenn nicht aufheben, so doch zumindest verzögern könnte.

Die fallende Grundfrequenz weiblicher Sprecherinnen ist ebenfalls so bereits in der Literatur zu finden (siehe hierzu ebenso Decoster und Debruyne (1997); Linville (1996, 2001); Baken (2005); Nishio und Niimi (2008); Harrington et al. (2007a)). Auch hier wirkt sich möglicherweise eine professionell trainierte Stimme (siehe in 2.3 die Daten aus Brown et al. (1991), die für Frauen keine Unterschiede zeigen) und andere Einflussfaktoren wie Rauchen, das bekannt ist, die Grundfrequenz abzusenken (Sorensen & Horii, 1982), auf die Ergebnisse aus (siehe auch in Abbildung 2.3 die nach Raucherinnen und Nichtraucherinnen getrennten Daten der Frauen von Linville), auch wenn die hier präsentierten Ergebnisse eher für ein „normales “ Altern der Stimme sprechen, d.h. die Grundfrequenz sinkt mit dem Alter. Im Vergleich hierzu fanden Brown et al. (1991) eben – wie schon beschrieben – keine altersbedingten Unterschiede bei professionellen Sängerinnen. Ein weiterer Unterschied zu den Befunden in der Literatur ist für die Daten der Königin auszumachen; während Linville (1996, 2001) davon ausgeht, dass es um ein Alter von 45 bis 55 bei Frauen zu einem menopausal hormonbedingten (Abitbol et al., 1999) plötzlichen Absacken der Grundfrequenz kommt (siehe auch Linvilles Daten in 2.3, und auch die longitudinalen Daten in De Pinto und Hollien (1982); Russell et al. (1995)), fällt *f0* der Königin eher stetig ab, was eher konsistent ist mit der Datensammlung aus Baken (2005) und den Daten aus Nishio und Niimi (2008).

Die zweite wichtige Erkenntnis in dieser Studie ist der Abfall des ersten Formanten bei allen Sprechern und nahezu allen Sprecherinnen (nur eine Sprecherin zeigt einen nur gering ausgeprägten Abfall des ersten Formanten). Auch dies ist vollständig konsistent mit der Befundlage in der Literatur (Linville & Fisher, 1985a, 1985b; Linville & Rens, 2001; Scukanec et al., 1991; Xue & Hao, 2003). Soweit es dem Autor bekannt ist, ist der Befund, dass bei Männern im höheren Alter *F1* auch wieder steigen kann, neu. Es ist in dieser Diskussion aufgrund der Datenlage in Teilexperiment 2.2.4 behauptet worden, dass *F1*

den Änderungen in *f0* folgt, was die dritte wichtige Erkenntniss aus dieser Studie ist. Wie man aus Experiment 2.2.3 ersieht, ist dies bei den beiden Sprechern Gillard und Muir, bei denen *f0* anstieg, nicht der Fall, denn bei diesen sinkt *F1* wie bei den anderen. Ist es also dennoch möglich, zu behaupten, dass *F1* und *f0* kovariieren?

Wie die Abbildungen 2.12 und 2.13 zeigen, ist auch bei einer beinahe jährlich vorgenommenen Reanalyse so viel Variation in den Daten enthalten, dass es durchaus der Fall sein kann, dass bei einer Zufallsauswahl aus diesen Aufnahmen (und auf solche Zufallsauswahlen beruhten die Teilexperimente 2.2.1 und 2.2.3) und der Verwendung von stimmhaften Signalanteilen der erste Formant sowie die Grundfrequenz mal höher und mal tiefer ausfallen, und zwar mit einer durchaus beachtlichen Variation. Über viele Jahre beobachtet (wie in Experiment 2.2.4), zeigt sich jedoch das in diesem Kapitel beschriebene Muster; die linearen Regressionen, die die Daten der Königin und Cookes modellieren, sind hinreichend valide, um uns über die Variation zwischen den Jahren hinwegsehen zu lassen. Bezogen auf die Sprecher Muir und Gillard, die das beschriebene Muster nicht aufweisen, kommt hinzu, dass sich für diese Sprecher die Materialmengen für *früh* und *spät* erheblich unterscheiden, und es hierbei wahrscheinlicher ist, dass dort nicht genügend Material in den frühen Aufnahmen vorhanden ist (jeweils circa drei Minuten), um phonetisch balancierte Sprache zu erwarten. Für die Formantwerte ist dort also – da ja nicht immer das gleiche Material gelesen wurde – mit einer größeren Anfälligkeit für zufällige Variation zu rechnen.

Weitere Hinweise auf eine mögliche Kovariation beider Parameter bietet Torre III und Barlow (2009), welche in einer Querschnittsstudie mit 27 jungen (circa Mitte 20) und 59 alten (circa Mitte 70) Sprecherinnen und Sprechern ebenfalls keine Beeinflussung von *F2* und *F3* finden, aber einen *f0*-Abfall bei den Frauen, kombiniert mit einer Absenkung des ersten Formanten, bei Männern jedoch eine steigende Grundfrequenz und –vokalabhängig – entweder steigende *F1*-Werte (bei /i,æ,ʊ/) bzw. nur leicht fallende *F1*-Werte (bei /ɪ,ɛ,ʌ/). Leider ist der dort angegebenen Statistik nicht zu entnehmen, ob es denn vokalübergreifend einen Haupteffekt des Alters auf den ersten Formanten bei Männern gab.

Gehen wir also – wie Reubold et al. (2010) – von einer altersbedingten Kovariation von *F1* und *f0* aus, so stellt sich die Frage nach den möglichen Gründen. Als erstes galt es, auszuschließen, dass es sich bei diesem Befund um ein bloßes Artefakt der LPC-Analyse handelte. In Teilexperiment 2.2.5 wurde durch Analyse von Resynthesen mit veränderten Grundfrequenzwerte eine gewisse Beeinflussung der *F1*-Werte gezeigt, die allerdings gänzlich anderer Natur und von anderem Ausmaß sind als die Befunde für die Königin und Cooke. Wir wollen also davon ausgehen, dass es sich *nicht* um ein Artefakt der Messung handelt. Allerdings wird Kapitel 3.3 zeigen, dass es gerade bei Frauenstimmen durchaus zu messungsbedingten scheinbaren Abhängigkeiten von *F1* und *f0* kommen kann. Diese sind dann allerdings so offensichtlich und vor allem systematisch, dass sie nicht schwer aufzudecken sind; sie beruhen in der Regel darauf, dass eine der Harmonischen die Formantmessung dergestalt beeinflusst, dass *F1* ein ganzzahliges Vielfaches von *f0* ist. Da ein solcher Einfluss hier auszuschließen ist, wollen wir die Hypothese, der Befund beruhe auf Artefakte, verwerfen.

Ein weiterer möglicher Grund liegt begründet in akustischer *F1–f0*-Interaktion. Eine Möglichkeit aus diesem Bereich liegt in der Kopplung sub- und supralaryngaler Räume.

Erstens gibt es nach Titze (2004) eine Beeinflussung der Glottalschwingung durch den ersten Formanten, d.h. die Effizienz der Energiewandlung von aerodynamischer zu akustischer Energie an der Glottis wird beeinflusst durch die Lage des ersten Formanten. Eine vergleichbare Beeinflussung (Ishizaka & Flanagan, 1972) wurde im Zusammenhang mit Intrinsischer Grundfrequenz diskutiert , nämlich dass in hohen Vokalen, bei denen *F1* und *f0* in relativer Nähe zueinander stehen, der Formant die Schwingung der Glottis beeinflusst und die Frequenz nach oben zieht. Dieses Kopplung wird in (Ewan & Ohala, 1979) und Ohala und Eukel (1987) als Quelle der intrinsischen Grundfrequenz verworfen: So treten Intrinsische Grundfrequenz-Effekte auch in Äußerungen nach dem Einatmen von Helium auf, obwohl dort zwar die Formanten, aber nicht die Grundfrequenz durch die unterschiedliche Schallgeschwindigkeit in dem anderen Medium beeinflusst werden, also sich Grundfrequenz und erster Formant gar nicht nahe kommen und sich somit also nicht beeinflussen können; desweiteren unterscheiden sich – entgegen der Hypothese von der akustischen Kopplung – die Grundfrequenzen in Nasalen, was eher auf den vokalischen Kontext zurückzuführen ist.

Ein zweiter Fall akustischer Kopplung ist folgender: die Öffnung an der Glottis beeinflusst die Ankopplung des subglottalen Raumes. Ist dieser angekoppelt, verschiebt die nun veränderte Impedanz der geöffneten Glottis die Frequenz des ersten Formanten, wobei er zusätzlich die Amplitude des ersten Formanten abschwächt (Holmes & Holmes, 2001; Klatt & Klatt, 1990; Childers & Wong, 1994). Barney et al. (2007) fand, dass der Open Quotient, also die relative Dauer der Öffnungsphase der Glottis während der Schwingungsperiode, und *F1* positiv korreliert sind, also *F1* steigt, wenn der Open Quotient steigt. Gleiches gilt, wenn grundsätzlich die durchschnittliche glottale Öffnung größer wird.

Daneben ist noch zu beachten, dass nicht alle Befunde der Literatur die gleichen Effekte vorhersagen: Berechnungen in Badin und Fant (1984) unter Benutzung von Formeln aus Flanagan (1965) ergaben weitaus geringer Effekte der glottalen Impedanz. Die Frage scheint also noch nicht abschließend geklärt zu sein.

Nehmen wir aber dennoch an, die Vorhersagen aus Barney et al. (2007) träfen zu, so müsste dies umgekehrt bedeuten, dass man, um das altersbedingte Sinken des ersten Formanten mit einer solcher Interaktion zu erklären, zeigen müsste, dass die glottale Öffnung sich mit dem Alter verringert oder der Open Quotient sinkt. In Teilexperiment 2.2.6 wurden diese Hypothesen mit akustischen Korrelaten dieser Maße überprüft – und die Hypothesen mussten verworfen werden. Zwar gibt es altersbedingte Veränderungen dieser glottalen Maße, aber jene wirken eher in die Gegenrichtung, d.h. *F1* müsste eher steigen, da im Durchschnitt sich die glottale Öffnung erweitert und auch der Open Quotient steigt. Wir wollen an dieser Stelle gar nicht ausschließen, dass dies einen *F1*-anhebenden Einfluss hat; im Endeffekt sinkt aber eben *F1*, so dass dieser Effekt – so er denn vorhanden ist – antagonistisch zu dem Effekt des noch unbekannten Faktors, der das Absinken bewirkt, arbeitet.

Desweiteren wäre natürlich zu fragen, wie eine solche Interaktion zwischen Glottis und Ansatzrohr dazu führen sollte, dass *F1* und *f0* in einer vergleichbaren Rate variieren.

Da diese – durchaus vorhandene – altersbedingte Variation der glottalen Öffnung und des Open Quotient also nicht ursächlich sein können für die altersbedingte *F1*-Variation, die in den Daten vorliegt, muss man nach Alternativerklärungen suchen. Eine davon beruht

auf physiologischen Veränderungen, die hypothetisiert werden und für Veränderungen der supralaryngalen Resonanzen verantwortlich sein könnten: die sogenannte *ptosis* des Kehlkopfs (Ferreri, 1959), also das Absinken des Kehlkopfs durch die altersbedingte Volumenänderung und daher Absenkung des Atmungs- und Verdauungstraktes sowie der Atrophie auch der extrinsischen Kehlkopfmuskulatur und durch zunehmende Dehnung der Bänder (Wilder, 1978; Laver & Trudgill, 1979; Kahane, 1980), die dann auch den Kehlkopf nach unten zieht und deshalb zu einem Absinken der Formanten führt, so wie (Linville & Rens, 2001) es beschreiben (wenn auch anhand der Gipfel in Long-Term Average Spectra, und dort am ausgeprägtesten beim ersten Gipfel).

Auch die Grundfrequenzvariation, die parallel zu derjenigen des ersten Formanten verläuft, wäre eventuell auf diese Weise erklärbar, da eine Korrelation zwischen Larynxhöhe und *f0* bekannt ist (Shipp, 1975; K. Honda et al., 1999). Larynxhöhe könnte ansatzweise auch erklären, warum *F1* und *f0* bei Cooke wieder ansteigen. Es gibt allerdings zwei Gegenargumente gegen diese Theorie. Der erste Einwand betrifft gerade den Wiederanstieg beim männlichen Sprecher Cooke. Wenn ein alterbedingtes generelles Absinken der Lunge, der Trachea usw. für die relative Lageveränderung des Kehlkopfes ursächlich sein sollte, stellt sich die Frage, weshalb im hohen Alter die relative Lage sich wieder nach oben verschieben sollte. Dies ließe sich eventuell allerdings erklären durch die Befunde in Iwarsson und Sundberg (1998), die – vielleicht überraschenderweise – eine negative Korrelation zwischen Lungenvolumen und vertikaler Kehlkopfposition feststellten, also z. B. eine höhere Lage des Kehlkopfs bei niedrigem Lungenvolumen. Zweifelsohne sinkt das Lungenvolumen mit zunehmendem Alter (Richards, 1965), insbesondere im hohen Alter – es wäre also vorstellbar, dass zuerst *Ptosis* den Kehlkopf nach unten „sacken“ lässt, und dieser im hohen Alter wegen der Effekte, die durch ein verringertes Lungenvolumen hervorgerufen werden, aktiv angehoben werden muss, um zu phonieren .

Ein weiterer Einfluss auf die Formanten könnte dann wieder mit den oben genannten altersbedingten Veränderungen der glottalen Öffnung zusammenhängen: Das Epithel der Stimmlippen von Männern wird dünner (Segre, 1971), weitere Volumenänderungenen treten in den verschiedenen Schichten der Stimmlippen auf (Sato & Hirano, 1997); zusätzlich ist bekannt, dass gerade bei Männern der Stimmlippenverschluss zunehmend nicht mehr vollständig ist (siehe den Überblick in Kapitel 3 bei Linville (2001)). Es wäre also denkbar, dass bei Cooke - oder bei Männern im Allgemeinen - eine Kombination von Änderungen der Lage des Kehlkopfs und innerhalb des Kehlkopfs zur erneuten Ansatzrohrverkürzung und gleichzeitig zu Veränderungen in der akustischen Kopplung der Glottisschwingung und des ersten Formanten zusammenwirken und so *F1* wieder ansteigen lassen. Weshalb dann allerdings diese deutliche Kovariation mit *f0* auftritt, die bestenfalls durch die Lageveränderung ansatzweise erklärt werden könnte, bleibt wieder fraglich.

Ein zweiter, wichtigerer Einwand gegen die Theorie der Altersabhängigkeit der Formanten von der Lage des Kehlkopfes ist genereller: Wie (Sundberg & Nordström, 1976) hypothetisieren, sollte eine Lageveränderung des Kehlkopfes generell das Ansatzrohr verkürzen bzw. verlängern und dementsprechend alle Formanten beeinflussen; ihre eigenen Messungen der Formanten bestätigen diese Hypothese. Lindblom und Sundberg (1971) modellierten den Vokaltrakt, um Formantänderungen zu berechnetn, die durch Änderun-

gen der Lippen und des Kiefers, der Form und Lage der Zunge sowie der vertikalen Position des Kehlkopfes hervorgerufen werden; sie stellten für eine Kehlkopfabsenkung ebenfalls ein Absinken aller Formanten fest - wenn auch in unterschiedlichen Ausmaßen: der ausgeprägteste senkende Einfluss (in %) wird allerdings auf *F2* ausgeübt, weniger Änderungen werden in *F1* hervorgerufen. Es ist also keineswegs klar, weshalb also eine altersbedingte Lageveränderung des Kehlkopfes nur (oder zumindest hauptsächlich) den ersten Formanten beeinflussen sollte, der ja nicht nur in dieser Studie der hauptsächlich, wenn nicht sogar ausschließlich konsistent betroffene Formant war (siehe den Überblick oben[11]).

Als Alternative ist also zu überlegen, ob auditorische Gründe eine Rolle spielen könnten; wie in der Einleitung ausführlich dargelegt, gibt es eine Reihe von Maßen, die sowohl *F1* als auch *f0* inkorporieren, und als eigentliche akustisch-auditive Korrelate von Vokal*höhe* (zumindest der Distinktion $[\pm high]$) in der Literatur genannt werden (Syrdal & Gopal, 1986; Traunmüller, 1981). Dieser Zusammenhang ist eng verwandt mit dem Befund, das bezüglich Vokalhöhe Grundfrequenz und erster Formant negativ miteinander korreliert sind, also mit steigender Zungenhöhe *f0* steigt (ein Effekt, der als Intrinsische Grundfrequenz bekannt geworden ist und wahrscheinlich universal in den Sprachen der Welt zu finden ist(siehe z. B. Whalen und Levitt (1995))), und *F1* fällt, so dass der Verdacht naheliegt, dass ein Maß, dass beide Parameter inkorporiert, erfolgreicher bezüglich Vokalhöhe Kategorien voneinander zu scheiden vermag. Ein weiterer Grundgedanke bei der Entwicklung dieser beide Parameter inkorporierender Maße ist der, dass die Anregung der Basilarmembran des Hörers eine Rolle spielt, was gleichzeitig auch eine Art der Sprechernormalisierung bedeutet; deshalb basieren die Maße bei Syrdal und Gopal (1986) und bei Traunmüller (1981) auf den Abstand zwischen *F1* und *f0* auf einer gehörgerechteren Skala, als die Frequenzskala in Hertz es ist, also z. B. auf der Bark-Skala.

Sollte also altersbedingt eines dieser Maße (*F1* oder *f0*) variieren, stellt sich die Frage, ob dann nicht auch die perzipierte Vokalhöhe davon beeinflusst wird. Sollte dies der Fall sein, ist vorstellbar, dass das andere Maß aktiv variiert wird, um das Vokalhöhenperzept wieder zu adjustieren. Die Ergebnisse für die langjährigen Verläufe von Cooke und der Königin, die in der der Bark-Skala nicht unähnlichen logarithmischen Skala gezeigt wurden, zeigen denn auch langfristig gesehen eine relativ stabil bleibende Distanz zwischen *F1* und *f0* – was dahingehend interpretiert werden könnte, dass beide Sprecher möglicherweise die Vokalhöhe stabil halten wollen.

Während es – wie in der Einleitung dargelegt – unklar ist, ob die weiter oben beschriebenen Effekte der altersbedingten Lageveränderung des Kehlkopfes denn tatsächlich nachzuweisen sind oder nicht – Flügel und Rohen (1991) und Xue und Hao (2003) waren nicht in der Lage, Kehlkopflageveränderungen bzw. Vokaltraktlängenunterschiede festzustellen, während die auch nach der Pubertät sich fortsetzende *Ptosis* des Kehlkopfes sogar in Lehrbüchern als Tatsachenbehauptung zu finden ist (Zemlin, 1998) – ist es unbestritten, dass es physiologische Änderungen im Kehlkopf und daraus resultierend verändertes Verhalten der Glottisschwingung gibt. Diese Veränderungen sind ausführlich in Linville (2001),

[11] Es soll nicht verschwiegen werden, dass für die Königin ein leichter Abfall des zweiten Formanten beobachtet wurde, der allerdings viel schwächer ausfiel als der des ersten Formanten; siehe hierzu 2.12

und zwar hauptsächlich im Kapitel 3, das sich ausschließlich mit Veränderungen des Kehlkopfes und seiner Funktion befasst, beschrieben. Neben diesen physiologischen Änderungen am Kehlkopf, die die Funktion der Glottis als Ventil negativ beeinflussen (Melcon et al., 1989; Hoit & Hixon, 1992) kommt es auch zu Veränderungen der Lungenfunktion und hier insbesondere der Sprechatmung (Hoit & Hixon, 1987; Huber & Spruill, 2008). Einer der Effekte all dieser Änderungen ist die altersbedingte Grundfrequenzvariation, wie sie auch in diesem Kapitel beschrieben wurde - also ein Abfall bei Frauen und ein mehr oder minder U-förmiger Verlauf, also ein Abfallen und anschließendes Ansteigen der *f0* bei Männern – beide Muster lassen sich mit den altersabhängigen physiologischen Änderungen in Einklang bringen. Sowohl bei Frauen wie Männern ist hierbei offenbar die Beeinflussung durch Sexualhormone stark (siehe z. B. Abitbol et al. (1999); Mendes-Laureano et al. (2006); Meurer, Osório Wender, von Eye Corleta und Capp (2004); Higgins und Saxman (1989); Schneider, van Trotsenburg, Hanke, Bigenzahn und Huber (2004) für Änderungen des Hormonspiegels und dessen Auswirkungen auf die Stimme bei Frauen und Gugatschka et al. (2010) für Änderungen bei Männern) und bedingt – unter Verringerung der morphologischen Unterschiede zwischen den Geschlechtern – z. B. zunächst eine fortschreitende Massenzunahme der Stimmlippen bei Frauen und - in sehr hohem Alter, eine Verringerung der Stimmlippenmasse bei Männern und sogar einer Verkürzung der Stimmlippen der Männer im Alter über 70 (Hirano et al., 1989).

Gehen wir also davon aus, dass die Veränderungen der Grundfrequenz, wie sie hier beschrieben wurden, unvermeidlich sind. Es stellt sich die Frage, ob diese Änderungen groß genug sind, um möglicherweise das Vokalhöhenperzept zu beeinflussen. Betrachten wir, dass bei der Königin ein maximaler altersbedingter Unterschied von fast 100 Hz (oder 6.8 Halbtönen) und bei Cooke von immerhin noch 35 Hz (oder 5.3 Halbtönen) auftritt, stellt man fest, dass dies in einer Größenordnung ist, die z. B. die Variation von Intrinsischer Grundfrequenz von /ɑː/ zu /iː/ in einer Person zu einem Zeitpunkt übersteigt (vergleiche z. B. die Werte in Katz und Assmann (2001), für Englisch, oder der sprachübergreifende Vergleich in Whalen und Levitt (1995), die Unterschiede zwischen den hohen Vokalen /iː,uː/ und dem tiefen Vokal /ɑː/ von 1.84 Halbtönen für Männer und von 1.34 Halbtönen für Frauen angeben). Nun ist Intrinsische Grundfrequenz so definiert, dass man ihre Effekte nur ausmachen kann, wenn man alles andere unverändert lässt („all other things being equal“ (Ohala, 1973)). Gehen wir von einem ähnlichen Grundsatz für den auditiven *F1-f0*-Abstand aus, und davon, dass – über alle Aufnahmen aus den verschiedenen Jahren betrachtet – auch die phonetischen Kontexte, die Phrasierungen usw., also alles was entweder Vokale und/oder die Grundfrequenz beeinflussen kann, in diesen zahlreichen Aufnahmen der Königin und Cookes mehr oder weniger ähnlich wie in der Grundgesamtheit der Äußerungen dieser Sprecher verteilt ist, besteht also durchaus der Verdacht, dass die beschriebenen altersbedingten *f0*-Änderungen das Vokalhöhenperzept beeinflussen müssen – und daher die Notwendigkeit entsteht, dafür kompensatorisch den ersten Formanten anzupassen.

Die Idee dahinter ist also vergleichbar mit dem, was Traunmüller (1991b, Seite 125) für den von ihm gefundenen Zusammenhang zwischen *F1* und *f0* beschreibt:

> Paradigmatic variation among speech sounds is, of course, essential for their distinctive linguistic functioning. At least for vowels, the formant frequencies carry most of that distinctive function. These formant frequencies are, however, also affected by paralinguistic and extralinguistic variation. Vowels have always such „personal quality“ in addition to their phonetic quality(...) . „Transmittal“ variation does not usually affect *f0* or *F1*. The covariation between *F1* and *f0*, which is our present concern, we can observe when we compare two linguistically identical utterances. If, in such utterances, *f0* is different due to any differences in personal quality, it is nearly always the case that also *F1* differs in the same direction.

Wie gesagt haben wir es bei den Daten der Königin und Cookes *nicht* mit linguistisch identischen Äußerungen zu tun, aber doch immerhin mit soviel Material, dass die schiere Masse es möglich macht, davon auszugehen, dass die Aufnahmen ähnlich vergleichbar sind wie identische Aufnahmen es wären. Ändert sich also *f0* altersbedingt, sollte irgendeine Form der kompensatorischen *F1*-Variation erfolgen. Eine solche kompensatorische *F1*-Veränderung ließe sich dadurch erreichen, dass man den Grad der Mundöffnung variiert. Als Lindblom und Sundberg (1971) den Vokaltrakt modellierten, stellten sie fest, dass über alle Kontexte hinweg der konsistenteste Effekt einer Zunahme der Kieferöffnung (und damit der Mundöffnung) eine Absenkung des ersten Formanten ist. Es ist also vorstellbar, dass Sprecher, bei denen sich über die Jahre hinweg *f0* ändert, darauf mit einer kompensatorischen Kieferöffnungsvariation reagieren. Tatsächlich wurden die auch hier vorliegenden Daten der Königin im Rahmen einer Sommerschule dazu benutzt, um ihren Vokaltrakt mittels Inversion (*acoustic-to-articulatory inversion*) zu modellieren, indem Maedas Modell (Maeda, 1979, 1990) angewendet wurde (Beyerlein et al., 2008). Hierbei wurde tatsächlich eine zunehmend angehobene Position des Kiefers für die Königin festgestellt. Leider ist dieses Resultat nicht sehr robust, da - wie persönliche Kommunikation mit Teilnehmern diese Sommerschule (Georg Stemmer, Elmar Nöth, und Anton Batliner) bestätigte, dieses Ergebnis insofern zirkulär zustandekommt, da die Kieferposition hauptsächlich über den ersten Formanten vorhergesagt wird, der – wie auch die vorliegende Studie zeigt – schließlich altersbedingt fällt.

Zusammenfassend ist also festzuhalten, dass altersbedingt Veränderungen der Grundfrequenz festzustellen sind; diese werden begleitet von Veränderungen des ersten Formanten; es ist nicht letztlich auszuschließen, erscheint aber unwahrscheinlich, dass die festgestellte Kovariation beider Parameter ein Zufallsprodukt ist, das zustandekam, da nur zwei Sprecher über viele Jahre hinweg nachverfolgt werden konnten. Eine weitere Möglichkeit ist, dass es zwar solche Variationen in beiden Parametern gibt, die aber keinen kausalen Zusammenhang haben. Die starke Kovariation beider Parameter – z. B. der Umstand, dass bei Cooke der Umkipppunkt für beide Parameter fast im gleichen Lebensjahr zu finden war – und die Steigungen der Variation, die einen mehr oder minder konstanten Abstand zwischen den Maßen vermuten lässt, lässt uns diese Möglichkeiten aber nicht für wahrscheinlich halten. Ein Artefakt scheinen wir ausgeschlossen zu haben. Festgestellte altersbedingte Änderungen der Öffnung an der Glottis können die gezeigten Änderungen

des ersten Formanten nicht erklären. Weitere physiologische Erklärungen für beide Effekte sind verfügbar, erklären aber letztendlich nicht, warum dieser Abstand so konstant zu sein scheint. Am wahrscheinlichsten erscheint dem Autoren die Hypothese, die am Ende dargelegt wurde: eine altersbedingte Variation der Grundfrequenz kann die Vokalhöhenperzeption beeinflussen, und veranlasst den Sprecher zur Kompensation im Bereich des Öffnungsgrades, dessen konsistentestes Korrelat der erste Formant ist.

Es bestehen eigentlich keine Möglichkeiten, diese Hypothese anhand der vorliegenden Aufnahmen alternder Sprecher zu falsifizieren zu versuchen. Stattdessen müssen wir in den folgenden Kapiteln feststellen, ob die Vokalhöhenperzeption durch die Grundfrequenz überhaupt beeinflusst werden kann, und ob wegen einer Perturbation des Vokalhöhenperzeptes nicht nur mit einer kompensatorischen Änderung des Öffnungsgrades, sondern auch mit Variierung der Grundfrequenz reagiert wird.

Kapitel 3

Perturbations-/Kompensationsexperimente

3.1 Einführung

In diesem zweiten experimentellen Kapitel wollen wir testen, inwieweit für eine Perturbation des Vokalhöhenperzeptes sowohl *F1* als auch *f0* für eine Kompensation eingesetzt werden kann, und inwieweit eine *f0*-Perturbation zu einer Wahrnehmung einer Vokalhöhenverschiebung, für die dann in *F1* kompensiert werden muss, führen kann. Wie wir im vorangegangenen Kapitel festgestellt haben, existiert eine relativ deutliche Kovariation zwischen beiden Parametern, wenn der Sprecher altert. Wie wir desweiteren festgestellt haben, lassen Befunde aus der Literatur es für sehr wahrscheinlich halten, dass die Grundfrequenz sich physiologisch bedingt mit dem Alter ändert; weniger klar ist, ob das Ansatzrohr und damit die Formanten sich physiologisch bedingt verändern. Die erwähnte Kovariation, welche zumindet für zwei Sprecher festgestellt wurde, legt den Verdacht nahe, dass diese mit einer angestrebten Invarianz der Vokalhöhe zu tun haben könnte, also dass die sich verändernde Grundfrequenz das Vokalhöhenperzept des Sprechers dergestalt perturbiert, dass irgendwann der Punkt erreicht ist, an dem der – dem Sprecher immer über das auditorische Feedback zugängliche – Vokalraum nach oben bzw. nach unten (je nach Richtung der Grundfrequenzänderung) verschoben erscheint. Hierfür – so unsere Hypothese – kompensiert der Sprecher nun mit einer Veränderung des Öffnungsgrades. Wir vermuten also hinter der *F1*-Kovariation mit *f0* einen vom Sprecher aktiv gesteuerten Prozess, der eine Rekalibrierung des Öffnungsgrades über den Abgleich mit dem auditorischen Feedback des Sprechers erfordert.

Die Rolle von Feedback in der Sprachproduktion Es erscheint offensichtlich, dass zur Kontrolle der Motorik oft eine Form von Feedback unerlässlich ist, zumal, wenn das zu erreichende Ziel außerhalb des eigenen Körpers liegt, beispielsweise bei dem Griff nach einer Tasse. Zwar kann man „blind“ nach etwas greifen, aber erst, nachdem man eine Repräsentation über den relativen Standort dieses Gegenstandes gebildet hat. Im Normalfall

sind wir jedoch auf das Feedback angewiesen, wie in diesem Fall des Greifens nach einer Tasse auf visuelles Feedback. Interessanterweise sind wir sehr schnell in der Lage, für Störungen oder *Perturbationen* des Feedbacks zu kompensieren und uns zu adaptieren, was schon Helmholtz Mitte des 19ten Jahrhunderts mittels der Störung des visuellen Feedbacks durch Prismen zeigen konnte, wobei die Forschung in diesem Bereich, beispielsweise anhand der Hand-Auge-Koordination, bis in die Gegenwart weitergeführt wird (siehe beispielsweise (Ghahramani, Wolpert & Jordan, 1996), denen auch der Hinweis auf Helmholtz' frühe Forschungen entnommen ist). Andererseits ist bekannt, dass manche Bewegungen schon „angeboren" sind; so zeigen überraschenderweise Blindgeborene mimische Bewegungen (Wolpert, Ghahramani & Flanagan, 2001); diese sind also offenbar nicht über visuellen Input erworben worden. Wo sind die artikulatorischen Gesten einzuordnen, d.h. wie viel Feedback ist zu ihrer Kontrolle nötig?

Während des Spracherwerbs ist noch keine voll ausgeprägte interne Repräsentation der artikulatorischen Gesten vorhanden, weshalb diese bei Kindern noch eine wesentlich höhere Variabilität aufweisen (Zharkova, Hewlett & Hardcastle, 2011). Durch das bis über die Pubertät hinausgehende Wachstum, dass mit teils drastischen Veränderungen vor sich geht, wie z. B. des zweiten Absinkens des Kehlkopfs bei Jungen in der Pubertät, ist über einen längeren Zeitraum keine stabile Umgebung gegeben (Fitch & Giedd, 1999), d.h. selbst eventuell bereits internalisierte Gestenmuster müssen in diesen Jahren ständig rekalibriert werden (Kent, 1976). Diese Rekalibrierung findet durch den Abgleich über somatosensorisches und auditorisches Feedback statt bis die (notwendigerweise schnellere) Feedforward-Kontrolle ausreichend trainiert ist, wobei die Details dazu noch weitgehend unbekannt sind (Smith, 2010)(siehe auch Guenther (1994) für eine Modellierung); das auditorische Feedback erlaubt hierbei den Abgleich mit den sprachlichen Äußerungen der anderen Sprecher, wobei vermutlich die sogenannten „Spiegelneuronen" die entscheidende Rolle spielen (siehe hierzu die Diskussion in Perkell (2010, Seite 9)). Es gibt Hinweise, dass selbst die 19-21-jährigen Versuchspersonen in Liu, Russo und Larson (2010), die neben Kindern (7-12) und älteren Erwachsenen (60-73) einer *f0*-Perturbation, also verändertem auditivem Feedback ausgesetzt waren, in ihrem Kompensationsverhalten – und damit möglicherweise in ihrem Nutzungsverhalten auditorischen Feedbacks – zwischen den Kindern und den älteren Erwachsenen rangieren, in dem sie zwar die gleichen Latenzzeiten wie ältere Erwachsene aufweisen (die geringer als bei den Kindern sind), aber dafür wie die Kinder in geringerem Ausmaß für die Perturbation kompensieren als die älteren Erwachsenen; d.h., dass die Lernphase für die Feedforward-Kontrolle möglicherweise bis ins junge Erwachsenenalter hineinreicht.

Einmal erlernt, scheint die Feedforward-Kontrolle auch relativ unabhängig von Feedback arbeiten zu können. Kelso und Tuller (1983) fanden keine signifikanten Vokalverschiebungen unter dem Einfluss medikamentös unterdrückter somatosensorischer Information und gleichzeitiger Verdeckung des auditorischen Feedbacks durch Rauschen, selbst dann nicht, wenn die Sprache durch einen Beißblock perturbiert wurde. Hoole (1987) hingegen fand bei einem Patienten, der krankheitsbedingt keinen Zugang zu afferenter somatosensorischer Referenz hatte, bei Beißblockexperimenten ohne und mit Verdeckung des auditorischen Feedbacks starke Unterschiede, was *für* eine wichtige Rolle des auditorischen Feed-

backs spricht, wenn eine ungewöhnliche Störung der Artikulation eintritt. J. A. Jones und Munhall (2003) benutzten eine Prothese, die die /s/-Produktion beeinträchtigte, und ließen ihre Versuchspersonen mit auditorischem Feedback eine neue Konfiguration erlernen, um danach das auditorische Feedback zu verdecken. In akustischen Messungen konnten sie zwar keine signifikanten Unterschiede zwischen den unverdeckten und verdeckten Produktionen feststellen, in Perzeptionsexperimenten wurden die unverdeckten aber besser bewertet; das auditorische Feedback wird also offenbar zur Feinjustierung auch gerade erlernter Gestensteuerungen immer noch benötigt. Ähnlich sind auch die Schlussfolgerungen in M. Honda, Fujino und Kaburagi (2002), die mittels einer dynamisch veränderbaren Gaumenprothese perturbieren und ebenso auditorischen Feedback unverdeckt und verdeckt anbieten; sie stellen zusätzlich fest, dass die Rekalibrierung über auditorisches Feedback offenbar langsamer ist als jene über die somatosensorische Schleife.

Man weiß, dass zumindest das auditorische Feedback ab einem bestimmten Alter an Wichtigkeit verliert, denn Menschen, die normalhörend den Spracherwerb vollzogen haben und erst spät ertauben, weisen, im Gegensatz zu gehörlos geborenen (Tye-Murray, 1987), sehr wenige Artikulationstörungen auf (Lane & Webster, 1991), die erst nach längerer Zeit zu entstehen scheinen, auch wenn die Grundfrequenz- und Intensitätskontrolle offenbar ein mehr an auditorischem Feedback verlangt, da sie bei spät Ertaubten am ehesten gestört wird (Lane & Webster, 1991; Lane et al., 1998) (siehe auch Perkell, Lane et al. (2007) für vergleichbare Effekte bei Trägern von Cochlea-Implantaten, mit ein- und ausgeschalteten Geräten). Lane et al. (1997) stellen die Theorie auf, dass die schnellen Änderungen phonetischer Parameter auf segmenteller Ebene bevorzugt eher *nicht* über die auditorische Feedback-Schleife kontrolliert werden, da die neuronale Verarbeitung über diesen Weg zu langsam ist, während auditorisches Feedback in ihrer Theorie eher zur Überwachung der akustischen Übertragung dient; verschlechtert sich die Übertragung, gibt es eher Änderungen in den suprasegmentalen Eigenschaften, weshalb diese bei Nicht-Gebrauch auditiven Feedbacks verstärkt werden. Lane et al. (1997, Seite 2251) halten hierzu fest: „When speakers think they may not be understood, one of the things they may do is to exaggerate the changes in prosodic parameters that they would normally make.“. Auch Perkell, Lane et al. (2007) haben ähnliche Befunde – nämlich geringe Auswirkungen auf segmentelle, aber starke auf suprasegmentelle – mit Cochlea-Implantat-Trägern als Versuchspersonen, denen unerwartet das Implantat ein- und ausgeschaltet wird, und schließen daraus: „...sound segment contrasts appear to be controlled differently from the postural parameters of speaking rate and average SPL and *f0*“(Perkell, Lane et al., 2007, Seite 2296).

Nasir und Ostry (2006) verweisen bezüglich der Rolle des Feedbacks auf segmentelle Eigenschaften auf die starke Rolle des somatosensorischen gegenüber des auditorischen Feedbacks, nachdem auch für leichte mechanische Perturbationen, die offenbar klein genug sind, um nicht zu auditorisch wahrnehmbaren Unterschieden zu führen, kompensiert wird.

McCaffrey und Sussman (1994) zeigen allerdings, dass Personen mit starken Hörverlusten hauptsächlich oberhalb des Frequenzbereichs des ersten Formanten eine höhere Variabilität entlang der [±*back*]-Achse haben, aber eine vergleichsweise „normale“ Variabilität in der Vokalhöhendimension; möglicherweise nutzen sie also doch noch den Rest des auditorischen Feedbacks im tieffrequenten Bereich, der ihnen noch zur Verfügung steht.

Auch die Ergebnisse aus Perkell, Denny et al. (2007) lassen auf eine andauernde Kontrolle der eigenen Produktionen auf segmenteller Ebene durch auditorisches Feedback schließen: Unter zunehmend durch Rauschen verdecktem Feedback verringert sich sehr schnell die Genauigkeit der Frikativproduktionen, und während bei geringer Verdeckung durch Rauschen der Vokalraum zunächst peripherer wird, verringert er sich bei zunehmender Verdeckung.

Perkell, Matthies et al. (2004); Perkell, Guenther et al. (2004); Villacorta (2006); Villacorta, Perkell und Guenther (2007); Brunner et al. (2011) zeigen, dass es einen Zusammenhang gibt zwischen der Feinheit der (auditiven) Perzeption und der Genauigkeit der Produktion, d.h. Hörer/Sprecher mit feinerer sensorischer Diskrimination bei der Perzeption erwerben schärfer voneinander abgegrenzte, distinkte Zielregionen in ihrer Feedforward-Kontrolle und erzeugen folglich weniger Überlappung bei der Produktion (Perkell, 2010) bzw. reagieren stärker auf Perturbationen des auditorischen Feedbacks (Villacorta, 2006; Villacorta et al., 2007). Ghosh et al. (2010) zeigen einen Zusammenhang sowohl zwischen auditorischer und somatosensorischer Sensibilität und der Genauigkeit von Frikativproduktionen.

Ein in den letzten zwei Jahrzehnten entwickelter Versuch, die hier genannten Befunde zu modellieren, stellt das hauptsächlich bei der Speech Communication Group des Massachusetts Institute of Technology entwickelte *DIVA*-Modell (*Directions Into Velocities of Articulators*) dar (Guenther, 1994, 1995; Guenther, Hampson & Johnson, 1998; Guenther, Ghosh, Nieto-Castanon & Tourville, 2006; Perkell, 2010). Da in diesem Kapitel kein Vergleich verschiedener Modellierungen angestrebt wird, soll nur dieses Modell, und dieses auch nur äußerst kurz dargestellt werden. Abbildung 3.1 zeigt das Modell als Schaubild; deutlich zu erkennen sind die zwei Hauptkomponenten: das *Feedforward Control Subsystem* und das *Feedback Control Subsystem*. Letzteres trainiert ersteres während des Spracherwerbs und wird nach Abschluss des Trainings zur Abstimmung und Feinjustierung genutzt. Es teilt sich in die somatosensorische und die auditorische Feedback-Schleife auf. Eintretende Abweichungen, die beispielsweise durch Perturbationen hervorgerufen werden, werden dort detektiert und veranlassen gegebenenfalls eine Rekalibrierung des *Feedforward Control Subsystems*. Dies ist immer dann der Fall, wenn die sogenannten *Goal Regions* nicht erreicht werden. Diese *Goal Regions* sind sprecher-/hörerabhängig; die Annahme ist, dass derjenige Hörer, der schlechter diskriminieren kann, auch größere *goal regions* hat, und deswegen weniger sensibel auf Perturbationen reagiert (siehe die obengenannten Ergebnisse aus Perkell, Matthies et al. (2004, 2004); Perkell (2010); Villacorta (2006); Villacorta et al. (2007); Ghosh et al. (2010); Brunner et al. (2011); man beachte aber auch, dass MacDonald, Purcell und Munhall (2011) keine Korrelation zwischen der Variabilität der Vokale „normaler“ Prduktion und dem Ausmaß der Kompensation für Formantperturbationen finden konnten). Desweiteren ist es der Fall, dass es beispielsweise im Fall akustischer Perturbation zu einem Konflikt zwischen der auditorischen und der somatosensorischen *Goal Region* kommen kann, und es mag sprecherabhängige Unterschiede im Ausmaß der Nutzung beider Domänen geben. Beide Effekte können dazu führen, dass Kompensationen für Perturbationen nicht vollständig sind, was häufig, wie zu sehen sein wird, der Fall ist. Interessant ist das Modell für dieses Kapitel aber auch deshalb, weil es einen weiteren Grund für unvollständige Kompensation für Perturbation geben könnte: die *Goal Regions* sind nicht a

priori eindeutig bestimmten (z. B. akustischen) Parametern zuzuordnen; sollte beispielsweise Vokalhöhe nicht allein über den ersten Formanten definiert sein, ist es vorstellbar, dass für Vokalhöhenperturbationen – seien sie nun mechanischer Natur wie in Beißblockexperimenten oder akustischer Natur wie in Formantperturbationen – nicht allein durch eine Korrektur des Öffnungsgrades kompensiert werden, sondern auch durch andere Parameter, wie die in diesem Kapitel zu untersuchende Grundfrequenz. Sollten sich hier systematische Effekte ergeben, kann man davon ausgehen, dass die Grundfrequenz ein entscheidender Faktor für die Erreichung der auditorischen *Goal Region* bezüglich der Vokalhöhe ist. Ein Vorteil wäre hier zu vermuten: Die – notwendigerweise entstehende – Diskrepanz zwischen der auditorischen und der somatosensorischen *Goal Region* wäre auf diese Weise vermutlich geringer, da eine geringere Öffnungsgradveränderung nötig wäre. Sollte sich diese Vermutung bewahrheiten, stellt sich die Frage, ob Perturbationen der Grundfrequenz „nur“ in der prosodischen Domäne wahrgenommen werden und dort zu Kompensationen führen, oder ob dadurch auch das Vokalperzept des Sprechers/Hörers beeinflusst wird, was sich durch Kompensationen in diesem Bereich auswirken sollte.

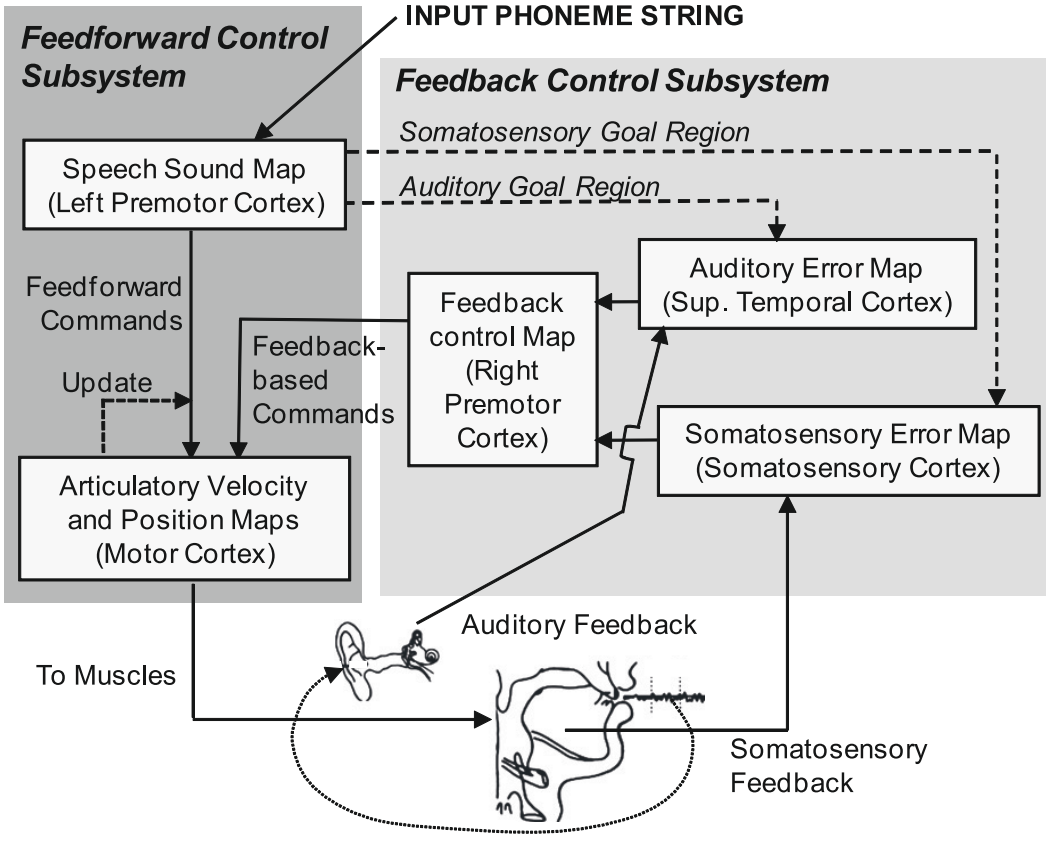

Fig. 1. A schematic block diagram of the DIVA model of speech motor planning and its interactions with an articulatory synthesizer.

Abbildung 3.1: *Das DIVA-Modell, zitiert nach Perkell (2010, Seite 3), mit originaler Abbildungsunterschrift.*

DIVA ist in der Hauptsache computational konzipiert und auf diese Weise getestet worden (worauf wir in dieser Arbeit nicht weiter eingehen werden, aber siehe hierzu z. B. Guenther, Ghosh und Tourville (2006)), ist aber auch schon in zahlreichen Perturbationsexperimenten hypothesenbasiert getestet worden, von denen einige im nachfolgenden Literaturüberblick genannt werden. Für jede Untersuchung des Zusammenhangs zwischen sensorischer Information und der Kontrolle von Bewegungsabläufen ist es sinnvoll, die Bewegungsabläufe selbst (z. B. durch Blockade der Kieferbewegungsmöglichkeit durch einen Beißblock) oder nur die sensorische Information (z. B. in dem man das akustische Signal, das dem Sprecher als auditorisches Feedback dient) zu perturbieren; man wird in diesen Fällen hypothetisieren, dass die Versuchspersonen für diese Perturbation auf eine bestimmte Weise und in einem bestimmten Ausmaß kompensieren; nach einer gewissen Zeit sollten die kompensatorischen Bewegungen erlernt sein, wobei man dann von sensorimotorischer Adaptation spricht; diese Adaptation wirkt sich so aus, dass eine Wegnahme der Perturbation zu den sogenannten *aftereffects* führt, da der Sprecher nun mit den momentanen Einstellungen wieder nicht die *Goal Regions* erreicht. Die *aftereffects* sind oft dergestalt, dass zunächst noch die neu erlernte, adaptierte Einstellung beibehalten wird (was einen Mismatch mit der *Goal Region hervorruft*), und dafür in Anschluss daran oft überschießend kompensiert wird, d.h. die ursprüngliche Einstellung der Artikulatoren wird nicht nur wieder erreicht, sondern in Richtung zur vorangegangenen Perturbation übertroffen. Wie gesagt, kann über Art und Ausmaß der Kompensation, Adaptation und der *aftereffects* darauf geschlossen werden, wie die Feedback-Schleifen genutzt werden, aber auch abgeschätzt werden, wie die zu erreichenden *Goal Regions* eigentlich definiert sind.

Mechanische Perturbation Da einige wichtige Arbeiten bereits oben erwähnt wurden, sollen hier nur die wichtigsten Techniken zusammengefasst werden, und nur eine kleine Sammlung von Beispielen gegeben werden; außerdem wollen wir in einem Exkurs auf das Zusammenspiel verschiedener Artikulatoren sowie auf den Einsatz der Grundfrequenz eingehen.

Es gibt mehrere wohletablierte Techniken, die Artikulatoren statisch oder dynamisch, sowie erwartet als auch unerwartet zu perturbieren; statische Perturbation bedeutet, dass man einen störenden Fremdkörper einbringt, oder eben nicht; es wird also perturbiert oder eben nicht; dies ist z. B. bei Beißblockperturbation der Fall wie z. B. in (Lindblom, Lubker & McAllister, 1977; Lindblom, Lubker & Gay, 1977; Kelso & Tuller, 1983; Hoole, 1987). Man findet in diesen Studien, sofern Feedback nicht auf die eine oder andere Weise gestört wird (siehe oben), eine erstaunliche Fähigkeit der Zunge, für die Kieferdislokation zu kompensieren, so dass nahezu die gleichen *Goal Regions* erreicht werden. Da diese Kompensation praktisch sofort einsetzt (Lindblom, Lubker & Gay, 1977) (zumindest was Formantfrequenzen von Vokalen angeht), also schon bevor auditorisches Feedback das Ohr des Sprechers erreichen kann, kann man davon ausgehen, dass hier die somatosensorische Information genutzt wird. Andererseits ist auch bei Benutzung eines Beißblocks, was erstens eine statische und zweitens dem Sprecher bewußte Perturbation ist, die Kompensation bei Frikativbildungen nicht sofort und nicht vollständig (Flege & Fletcher, 1988).

Neben dem Beißblock gibt es noch andere Methoden, die Kieferbewegung zu stören, die auch erlauben, die Perturbation unerwartet einsetzten zu lassen, so z. B. durch eine mechanische Vorrichtung, die eine bestimmte Krafteinwirkung entgegen der schließenden Kiefergeste ermöglicht, wie in (Folkins & Abbs, 1975; Kelso, Tuller, Vatikiotis-Bateson & Fowler, 1984).

Einige Arbeiten beschäftigten sich mit mechanischer Perturbation der Unterlippe durch eine unerwartet einsetzende störende Krafteinwirkung, so z. B. Abbs und Gracco (1984), der sofort einsetzende korrigierende Bewegungen der Unterlippe, aber auch kompensatorische Bewegungsverstärkung der Unterlippe und anderer Gesichtsmuskeln zeigt, oder Munhall, Lofqvist und Kelso (1994), der für /i'pip/-Äußerungen Kompensationen in der Unterlippe, der Oberlippe und dem Kiefer zeigt, die aber *nicht* sofort einsetzen, aber auch eine Verzögerung der Abduktion der Larynx (wegen der „Verlängerung“ des vorangegangenen Vokals) sowie eine verlängerte Adduktionsphase, was ebenso für eine Kopplung oraler und laryngaler Artikulation spricht.

Künstliche Gaumenplatten, die hauptsächlich die Bildung koronaler Frikative beeinträchtigen, können statisch (McFarland, Baum & Chabot, 1996), aber auch dynamisch zum Einsatz kommen, letzteres z. B. in M. Honda et al. (2002) . Hierbei wird ein künstlicher Gaumen eingesetzt, der durch eine aufblasbare Kammer im Volumen veränderbar ist, was sowohl unerwartete statische als auch während der Artikulation sich verändernde, also dynamische Perturbation, erlaubt.

J. A. Jones und Munhall (2003) setzen eine Zahnprothese ein, die die Länge der Schneidezähne statisch verändert, was sowohl die aerodynamischen Bedingungen zur /s/-Bildung als auch die Akustik verändert.

Ein weiteres Instrument perturbiert nicht nur die Artikulation, sondern verändert auch ohne signalverarbeitende Mittel die resultierende Akustik; die Rede ist von einer Lippenröhre, die zwischen den Lippen gehalten werden muss, während gerundete Vokale geäußert werden (Savariaux, Perrier & Ohteru, 1995; Ménard, Perrier & Savariaux, 2004); die Länge der Röhre wird so gewählt, dass das Ansatzrohr gegenüber einer natürlichen Produktion gerundeter Vokale nicht künstlich verlängert wird (z. B. 20-25 mm für Erwachsene); der Durchmesser ist natürlich unveränderlich, und beträgt für Erwachsene 25 mm (in Savariaux et al. (1995) mit 20 mm angegeben; die Autoren berichtigten diese Angabe später). Um gleiche Bedingungen für Produktionen mit und ohne Lippenröhre zu schaffen, wird ein (kleiner) Beißblock verwendet. Zur eigentlichen Perturbation, die an dieser Stelle interessiert, wird die Röhre eingebracht, die die Querschnittsfläche an dieser Stelle des Ansatzrohrs vergrößert und somit Veränderungen des akustischen outputs hervorruft; für /u:/ werden die Formanten, insbesondere aber *F2*, angehoben (Details in Savariaux et al. (1995)). Bei Savariaux et al. (1995) wird mittels Teleradiographie, einer besonderen Art des Einsatzes von Röntgenstrahlen, die Konfiguration des Ansatzrohrs gemessen. Sie stellen starke sprecherspezifische Unterschiede in der Art und dem Ausmaß der Kompensation fest. Sprecher, die besonders gut das akustische Signal an ihre unperturbierten /u:/-Äußerungen anpassen können, erreichen dies durch eine Rückverlagerung der Konstriktion von einem velopalatalen zu einem eher velo-pharyngalen Bereich, wenn dieser auch nur von einer eizigen der elf Versuchspersonen tatsächlich erreicht wird (diese Person hatte auch dies besten

akustischen Kompensationen im *F1-F2*-Raum aufzuweisen).

Exkurs: Zusammenspiel mehrerer artikulatorischer Parameter Die Ergebnisse aus Savariaux et al. (1995) sind vergleichbar mit Befunden aus artikulatorischen Studien (ohne Perturbation, mit Ausnahme von Riordan (1977)), die zeigten, dass für die (durchaus zu Recht) sogenannten *gerundeten* Vokale nicht alleine die Lippenrundung entscheidend ist. Riordan (1977) und Hoole und Kroos (1998) zeigen eine Kovariation von Lippenrundung und vertikaler Kehlkopfposition, die sich gegenseitig zur Verlängerung des Vokaltraktes für gerundete Vokale ergänzen; wird die Lippenvorstülpung perturbiert, verstärkt sich kompensatorisch die Kehlkopfabsenkung Riordan (1977) (doch siehe auch die Modellierung der Manöver für den /i-y/-Kontrast in Wood (1986); dort wird zwar auch eine Involvierung der vertikalen Kehlkopfposition behauptet, eine kompensatorische Funktion des Kehlkopfes für labialen *undershoot* jedoch abgelehnt). Perkell, Matthies, Svirsky und Jordan (1993) stellen – neben der Lippenrundung – eine Anhebung des Zungenkörpers bei [u]-Produktionen fest; sie argumentieren, dass diese Anhebung eine velo-palatale Konstriktion bildet, die ähnliche akustische Auswirkungen wie die Lippenrundung habe, in der Hauptsache eine Absenkung des zweiten Formanten. (de Jong, 1997) untersucht die gleiche Fragestellung und bekommen, wie die Vorgängerstudie, relativ gemischte Resultate, was für sprecherabhängiges Verhalten spricht. Tabain und Perrier (2007) vermuten eher eine die Lippenrundung unterstützende Rückverlagerung des Zungendorsums, was wiederum konform geht mit Savariaux et al. (1995), sowie eine Retraktion der Zungenspitze. *Gerundete* Vokale zeichnen sich also durch ein Zusammenspiel vieler Artikulatoren aus, die sprecherabhängig in unterschiedlichem Ausmaß genutzt werden. Schlagwortartig zusammengefasst werden solche Effekte als „motor equivalence“, also „the capacity of a motor system to achieve the same end-product with considerable variation in the individual components that contribute to that output“ Hughes und Abbs (1976, Seite 199).

Die ansatzrohrverlängernde Wirkung einer Kehlkopfabsenkung ist möglicherweise nicht die einzige laryngale Komponente während der Produktion gerundeter Vokale. Savariaux, Perrier, Orliaguet und Schwartz (1999) führen Perzeptionsexperimente mit den aus Savariaux et al. (1995) gewonnenen akustischen Aufnahmen unperturbierter und perturbierter /uː/-Produktionen und finden, dass die /uː/-Äußerungen scheinbar schlecht kompensierender Versuchspersonen (schlecht in der Domäne der bislang betrachteten artikulatorischen Zungendaten und des *F1-F2*-Vokalraums) besser bewertet werden als erwartet, und die Äußerungen z. B. des scheinbar optimal durch Rückverlagerung der Zunge kompensierenden Sprechers als weniger gut als erwartet. Eine Analyse der akustischen Korrelate der perzeptuellen Ergebnisse offenbart eine offensichtliche Involvierung der Grundfrequenz bei der Bewertung. Die besten Korrelate scheinen in der Abbildung innerhalb einer $[F1, (F2 - f0)]$-Ebene zu finden zu sein, und der Parameter $(((F2 - F0) + F1)/2)$ scheint das beste Korrelat für das phonologische feature *grave*, das für /uː/ sprachübergreifend relevant ist, darzustellen. Ähnliche Befunde liefert die bereits oben erwähnte Lippenröhrenperturbationsstudie in Ménard et al. (2004), die Kinder als Versuchspersonen verwendet, und ebenso starke sprecherabhängige Kompensationen findet. Beide Studien weisen also auf sprecherspezifische

artikulatorische Unterschiede unter Nutzung relativ invarianten auditorischen Feedbacks, welches offenbar die Grundfrequenz als eines der akustischen Eigenschaften mit einschließt.

Auch Mooshammer et al. (2001) zeigen eine Involvierung laryngaler Parameter bei Beißblockexperimenten: bei einer Perturbation der Artikulation durch einen Beißblock wird die vokal-intrinsische Grundfrequenz in verstärktem Maße eingesetzt, was die Autoren so deuten, dass die Intrinsische Grundfrequenz wohl auch aktiv zur Kompensation eingesetzt werden könnte (auch wenn sie diese Interpretation eher aus den unterschiedlichen Zungenhöhen bei vergleichbarer Grundfrequenz in *gespannt-ungespannt*-Paaren ziehen). Man kann diesen Befund aber auch rein mechanisch motiviert beschreiben: Ohala und Eukel (1987) erhalten ebenfalls bei einem Beißblockexperiment das Ergebnis, dass die Grundfrequenzvariation in Abhängigkeit von Vokalklassen sich unter Beißblockperturbation verstärkt, deuten dies jedoch als Evidenz für die sogenannte tongue-pull-hypothesis, die von einer rein mechanischen Kopplung zwischen Zunge und Kehlkopf ausgeht. Da in beiden Fällen keine Elektromyographie-Daten für Kehlkopfmuskeln vorliegen, muss der Grund für die Erweiterung der vokalintrinsischen F0-Variation im Unklaren bleiben. Lubker, McAllister und Lindblom (1977) finden bei schwedischen Sprechern keine vergleichbaren Effekte der Beißblock-Bedingung auf die intrinsische Grundfequenz, sondern sogar *tiefere f0*-Werte unter Perturbation.

Akustische Perturbation Bislang durchgeführte Perturbationen des auditorischen Feedbacks durch Manipulation des akustischen Signals beschränken sich zumeist auf Grundfrequenz oder Formanten, wenn man die globaleren Manipulationen durch Verzögerungen (*delayed auditory feedback*, Fairbanks (1955); Yates (1963)) und Wiedergabelautstärke und/oder durch Zumischung von Rauschen, um den wohlbekannten Lombard-Effekt hervorzurufen (Lane & Tranel, 1971; Siegel, Pick, Olsen & Sawin, 1976; Junqua, 1996), einmal außer acht lässt (siehe auch Liu, Zhang, Xu und Larson (2007) für lokale Lautheitsperturbationen, die dann als prosodische Störungen, nämlich als Fokusverschiebung, wahrgenommen werden); erst in jüngster Zeit gibt es Ansätze, auch das auditorische Feedback von Konsonanten zu stören, bislang jedoch eigentlich nur die spektralen Eigenschaften von koronalen Frikativen (Shiller, Sato, Gracco & Baum, 2009; Casserly, 2011). Grundsätzlich ist der Aufbau bei diesen Experimenten vergleichsweise ähnlich: die Versuchspersonen sprechen (oder singen) in ein Mikrophon; das Signal wird in ein Gerät geleitet, das die gewünschte Perturbation vornimmt und die Sprecher bekommen, zumeist über Kopfhörer, ihr eigenes Sprachsignal zu hören; zumeist gibt es verschiedene Ansätze, andere Wege des auditiven Feedbacks zu unterdrücken; in der Regel wird das Feedback über Kopfhörer relativ laut wiedergegeben, teilweise unter Benutzung von In-Ohr-Kopfhörer, die den Ohrkanal abdichten und so die gewöhnliche Übertragung über die Luft ausschließen sollen; ein weiterer, natürlicher Weg des Feedbacks ist die Übertragung durch die Schädelknochen, deren Einfluss gelegentlich durch eine Kombination lauten Feedbacks mit Überlagerung von (zumeist pinkem) Rauschen über die Kopfhörerbahn minimiert werden soll.

Perturbation der Grundfrequenz Seitdem die Methoden digitaler Signalverarbeitung so weit fortgeschritten waren, dass es möglich wurde, bestimmte einzelne akustische Parameter ohne wahrnehmbaren Zeitverzug zu manipulieren, wurden diese Mittel auch ausgeschöpft. Kawahara (1993) perturbierte die Grundfrequenz während eines ausgehaltenen /aː/-Vokals, worauf die Versuchspersonen mit einer kompensatorischen Veränderung der produzierten Grundfrequenz reagierten, bei allerdings nicht unerheblichen Latenzzeiten von 100-200 ms. Auch Burnett, Freedland, Larson und Hain (1998) und Max, Wallace und Vincent (2003) fanden vergleichbare kompensatorische Grundfrequenzbewegungen, und Larson, Burnett, Kiran und Hain (2000) stellten fest, indem sie die *f0*-Perturbation mit Rampenphasen unterschiedlicher Dauer einführten, dass für eine „sanft“ eingeleitete Perturbation schneller und vollständiger kompensiert wird. J. A. Jones und Munhall (2000) stellen negative *aftereffects* fest, d.h. nach dem Ausschalten einer Perturbation bleibt die Grundfrequenz zunächst in Gegenrichtung zur Perturbation verschoben; die Sprecher müssen sich also erst deadaptieren. Hawco, Jones, Ferretti und Keough (2009) stellen für 100ms-Perturbationen in 3 Sekunden ausgehaltenen Vokalen eine Mismatch-Negativity und eine schnelle Kompensation für die kurzen Perturbationen fest, die proportional umso geringer ausfallen, je stärker die Verschiebung ist (bei Hawco et al. (2009) bis zu zwei Halbtöne). Einige Studien beschäftigten sich mit Grundfrequenzperturbationen bei Sprechern von Tonsprachen, in denen also der Tonverlauf lexikalische Bedeutung überträgt: Xu, Larson, Bauer und Hain (2004) fanden gegenüber statischen Tönen bei dynamischen Tönen schnellere und stärker ausgeprägte Kompensation für Perturbation. J. A. Jones und Munhall (2005) stellen für Mandarin-Sprecher ebenso wie für die amerikanischen Sprecher in J. A. Jones und Munhall (2000) *aftereffects* fest, und dass diese Effekte auch in einer anderen Tonkategorie, die während der Perturbationsphase nicht geäußert worden war, beibehalten werden. Besonders interessant ist der Umstand, dass die Mandarin-Sprecher stärker für Grundfrequenzperturbationen kompensieren (circa 40%) als die amerikanischen Sprecher in J. A. Jones und Munhall (2000) (circa 25%).

Grundsätzlich ist das Ausmaß der Kompensation für *f0*-Perturbationen stark von der Aufgabe der Sprecher abhängig. So ist die Kompensation größer, wenn die Aufgabe eher an Singen erinnert, als in reinen Sprechaufgaben (Natke, Donath & Kalveram, 2003) (66% bei Singen, 47% bei Sprechen). In einzelnen Silben oder gar einzelnen Vokalen ist dennoch – wohl auch wegen der großen Latenzzeit – die Kompensation geringer als in Phrasen (Chen, Liu, Xu & Larson, 2007). Auch Donath, Natke und Kalveram (2002) hatten bereits argumentiert, dass das Ziel der Kompensation eher die Erhaltung suprasegmentaler Eigenschaften, und nicht das Erreichen eines pitch-targets innerhalb eines Silbenkerns ist.

Desweiteren ist es generalisierbar über alle Studien, dass es eine starke Sprechervariation des Ausmaßes der Kompensation gibt. Während manche Sprecher annähernd komplett kompensieren, kompensiert die Mehrheit nur unvollständig, und einige folgen sogar der Verschiebung, produzieren also z. B. bei nach oben verschobener Grundfrequenz höhere *f0* als in der sogenannten *Baseline*, also während jener Phase ohne Perturbation. Man findet auch Variation zwischen unterschiedlichen Sprecher- bzw. Sängergruppen, wie z. B. in J. A. Jones und Keough (2008), die einen Einfluss des Geübtseins auf den Grad der Kompensation feststellten: geübte Sänger kompensierten später als Nicht-Sänger, und benötigten im Ge-

gensatz zu Nicht-Sängern eine Deadaptationsphase (sie zeigten also *aftereffects*), was für eine stärkere Nutzung von Feedforward-Kontrolle bei geübten Sängern spricht [1].

Teilweise wird auch ein genereller Trend zur Anhebung der produzierten Grundfrequenz im Verlauf des Experiments festgestellt, so z. B. in J. A. Jones und Munhall (2000); dies kann mit dem Versuchsaufbau an sich zu tun haben (vergleiche z. B. den ebenfalls gefundenen generellen *f0*-Anstieg in Max et al. (2003) oder sogar im Formantperturbationsexperiment in Villacorta (2006)), aber auch mit der dem Sprecher möglichen Spannweite an *f0*-Produktion. Wenn der Sprecher in der *Baseline* bereits am unteren Ende seiner *f0*-Spannweite spricht, wird er Schwierigkeiten haben, als Kompensation zu einer nach oben gerichteten Perturbation der Grundfrequenz noch tiefere *f0* zu produzieren. Wenn man, wie J. A. Jones und Munhall (2000) den generellen Anstieg der Grundfrequenz herausrechnet (sofern ein solcher Anstieg überhaupt zu finden ist), ergeben sich in aller Regel keine Kompensationsunterschiede mehr, die der Perturbationsrichtung geschuldet sind.

Perturbation von Formanten Houde und Jordan (1998, 2002) ließen Versuchspersonen /CVC/-Wörter flüstern, um ihnen ohne wahrnehmbaren Zeitverzug formantverschobenes Feedback über Kopfhörer zu präsentieren, wobei die ersten drei Formanten verschoben wurden, so dass /ɛ/ in Richtung von /i/ verschoben erschien. Geflüstere Sprache wurde gewählt, um das Problem des Feedbacks durch die Schädelknochen zu minimieren. Es wurde eine Rampenphase benutzt, die Perturbation also allmählich eingeführt, und die Sitzungen waren von erheblicher Dauer von mehr als einer Stunde. Es wurde Kompensation festgestellt, also eine Gegenbewegung der produzierten Formanten zur Perturbationsrichtung. In Houde und Jordan (2002) wurde nach einer langen Phase unter Perturbation das Feedback durch reines Rauschen ersetzt, das jegliches auditorischen Feedback unterdrücken sollte. Die produzierten Formanten änderten sich wenig, was als Adaptation gedeutet wurde. Diese Adaptation wurde von den Sprechern auch auf den gleichen Vokal in anderen Kontexten als in der Perturbationsphase übernommen (z. B. /gɛg/ statt /pɛp/), und sogar auf andere Vokale (z. B. /pip/ statt /pɛp/) generalisiert. Allgemein wurde festgestellt, dass manche Versuchspersonen fast vollständig, andere gar nicht kompensieren (was für fast alle der hier genannten Studien gesagt werden kann).

Purcell und Munhall (2006a) manipulierten nur einen einzelnen Formanten, nämlich *F1*, in stimmhafter Sprache in Form von /CVC/-Wörtern. Es wurden Kompensationen gefunden, solange *F1* um mehr als eine bestimmte Schwelle (60 Hz) perturbiert wurde. Adaptation wurde festgestellt, d.h. es gab *aftereffects* nach der plötzlichen Rückkehr zu unperturbiertem Feedback; das Ausmaß der Kompensation und der Adaptation war nicht beeinflusst von der Perturbationsrichtung, und die Geschwindigkeit der Deadaptation unter unperturbiertem Feedback war nicht beeinflusst von der Dauer der vorangegangenen Perturbationsphase. Versuchspersonen mit größerer Variablilität in der *Baseline*-Phase ohne Perturbation zeigten unter der Perturbationsbedingung geringere Kompensation.

Purcell und Munhall (2006b) ermittelten in einem Vortest die mittlere Lage des ersten

[1] Geübte Sänger sind häufiger der Situation ausgesetzt, singen zu müssen, wenn sie sich selbst wegen anderer Mitsänger/ -instrumentalisten schlecht hören.

Formanten in /ɪ/ und /æ/, sowie des dazwischen liegenden /ɛ/, und ließen im eigentlichen Experiment Sprecher den Vokal /ɛ/ sprechen. Sie perturbierten diese Äußerungen unerwartet mit *F1*-Verschiebungen zu einem der beiden angrenzenden Vokale. Die gefundenen Kompensationen betrugen im Mittel 16.3% (Perturbation in Richtung /æ/) bzw. 10.6% (in Richtung /ɪ/); der interessanteste Befund ist wahrscheinlich der, dass die Latenzzeit für die Kompensation auf unerwartete *F1*-Perturbation nicht viel größer war, als in Grundfrequenzperturbationsexperimenten, nämlich circa 460 ms. Dies spricht für eine vergleichbare Verarbeitungsstrategie.

MacDonald, Goldberg und Munhall (2010) perturbierten *F1* und *F2* und verglichen, ob ein Unterschied besteht, wenn die Formanten sukzessive perturbiert werden (also mit einer sogenannten *Rampe*) oder ob die Perturbation, in verschiedenen Perturbationsstärken, plötzlich einsetzt, also in diesem Fall stufenweise. Die Resultate lassen vermuten, dass es keinen Unterschied macht, ob man eine Perturbation einer bestimmten Perturbationsstärke mit einer Rampe einführt oder plötzlich einsetzen lässt; allerdings ist das Ausmaß der Kompensation von der Perturbationsstärke abhängig, da bei Formantverschiebungen von mehr als circa 200 Hz (bei *F1*) bzw. 250 Hz (bei *F2*) ein Plateau in der kompensatorischen Antwort erreicht wird. Vor diesen Schwellwerten wird partiell kompensiert, und zwar relativ konsistent um circa 25-30% der angewendeten Perturbation.

Cai, Boucek, Ghosh, Guenther und Perkell (2008) und Cai, Ghosh, Guenther und Perkell (2010) zeigten, dass Kompensationen für Formant-Perturbationen nicht allein auf quasistatische Monophtonge beschränkt ist, sondern dass auch für dynamische Perturbationen im Mandarin-Triphtong /iau/ partiell kompensiert wird.

Munhall, MacDonald, Byrne und Johnsrude (2009) manipulierte die ersten beiden Formanten in (amerikanisch-englischen) *head*-Äußerungen, so dass das auditorische Feedback in Richtung *had* verschoben erschien; das Experiment wurde drei mal durchgeführt, und zwar erstens ohne die Versuchspersonen von einer Manipulation der eigenen Stimme zu informieren, zweitens indem man die Sprecher davon informierte, ihre Stimme könne anders klingen und deshalb sei das Feedback über Kopfhörer zu ignorieren, und drittens indem die Versuchspersonen explizit darauf hingewiesen wurden, dass ihre *head*-Äußerungen wie *had* klingen würden, was aber zu ignorieren sei; man solle auf keinen Fall dafür kompensieren. Erstaunlicherweise gab es keine nennenswerten Unterschiede der Antworten in den drei Bedingungen, d.h. in allen drei Fällen wurde in etwa gleich viel kompensiert, was dafür spricht, dass Kompensation für Perturbation ein automatischer Prozess und nicht vom Sprecher bewußt gesteuert ist.

Max et al. (2003) führten neben den oben erwähnten Grundfrequenzperturbationen auch Formantperturbationsexperimente durch, wobei während dieser Perturbation, bei der mittels eines kommerziellen Systems alle Formanten in *eine* Richtung verschoben wurden, artikulatorische Daten mittels Elektromagnetische Midsagittaler Artikulographie (EMMA) aufgezeichnet wurden. Sie stellten zwar für die Perturbation (im Gegensatz zu Houde und Jordan (1998, 2002) auf stimmhafte Äußerungen angewendet) relativ konsistente Kompensationen (und auch *aftereffects*) in der akustischen Domäne fest, aber auch starke Inkonsistenzen auf artikulatorischer Ebene, und zwar in der Hauptsache in Form von Zwischen-Subjekt-Variabilität, was für den Gedanken der bereits oben erwähnten *motor equivalence*

spricht; d.h. das gleiche Ziel (Wiederherstellung eines ähnliches auditorischen Feedbacks) kann auf unterschiedliche Weise artikulatorisch erreicht werden.

Noch ist die Datenlage etwas unklar über die Frage, ob auf eine Perturbation nur eines Formanten so kompensiert werden kann, dass nur in dem betroffenen Formanten kompensatorische Gegenbewegungen zu finden sind, oder ob nicht notwendigerweise die Kompensation multidimensional erfolgen muss, schon allein aufgrund artikulatorischer Beschränkungen. Katseff, Houde und Johnson (2010) findet denn auch bei Perturbationen einzelner Formanten Änderungen der Produktion sowohl von *F1* als auch *F2*. Im Gegensatz hierzu finden MacDonald et al. (2011) durch Experimente, in denen entweder *F1* oder *F2* perturbiert worden war, keine Änderung der Produktion im jeweils anderen Formanten, und befinden, dass Sprecher prinzipiell durchaus zu unabhängiger Kontrolle einzelner Formanten in der Lage sind.

Villacorta (2006) und Villacorta et al. (2007) zeigen, neben dem bereits erwähnten Zusammenhang zwischen der Feinheit der Perzeption eines Individuums und des Ausmaßes seiner Kompensation eine Adaptation der Sprecher, die bei Verdeckung des Feedbacks durch Rauschen auf andere konsonantische Kontexte und andere Vokale in /CVC/-Wörtern generalisiert wurde; am wichtigsten für die vorliegende Arbeit ist jedoch der Befund einer multidimensionalen Antwort: auf *F1*-Perturbatione reagieren die Sprecher nicht nur mit einer unvollständigen Gegenbewegung im ersten Formanten, sondern auch mit einer eher gering ausfallendes systematischen Variation in *F2*, und einer recht deutlichen systematischen Variation in *f0*, die *in Richtung* der *F1*-Perturbation (oder in anderen Worten: *in Gegenrichtung* zur produzierten *F1*-Verschiebung der Sprecher) stattfand („It was found that subjects modified F0 in a direction opposite to the compensatory F1 shift they produced“ (Villacorta et al., 2007, Seite 2311)). Die auditorische *Goal Region* ist laut der Autoren also charakterisiert durch Dimensionen, die von mehreren Formanten und der Grundfrequenz gebildet werden.

Beispiele für „natürliche Perturbation“. Unter Perturbation versteht man gewöhnlich eine von außen eingebrachte Störung, wie z. B. einen Beißblock, der der Beweglichkeit des Kiefers beeinträchtigt, wie z. B. in Hoole (1987), oder eine mechanische Störung der Unterlippe, wie in Munhall et al. (1994), oder die eben beschriebenen Perturbationen des auditorischen Feedbacks. Die Autoren des letztgenannten Papers (Munhall et al., 1994) gehen ausdrücklich davon aus, dass solche Störungen in „normaler“ Sprachproduktion äußerst selten vorkommen, und lassen als Fall einer „natürlichen“ Perturbation nur Fälle gelten wie den von ihnen genannten der Sopranistin Eva Marton, die auf der Opernbühne einen Unfall erlitt, der die Beweglichkeit ihres Kiefers stark beeinträchtigte, was sie aber durch kompensatorischen Einsatz der Zunge ausgleichen konnte (Munhall et al., 1994, Fußnote 1 auf Seite 3615). Andere, vielleicht weniger „natürliche“, aber doch „alltägliche“ Perturbationen wären beispielsweise Veränderungen am Gebiss, so z. B. Zahnspangen oder Zahnersatz, insbesondere wenn dieser als Prothese mit einer zusätzlichen Gaumenplatte getragen werden muss, was die Frikativbildung, insbesondere bei koronalen Frikativen, erheblich perturbieren kann (vergleiche die Perturbationsstudien in McFarland et al. (1996), M. Honda et al.

(2002) oder J. A. Jones und Munhall (2003)).

Andere Autoren weisen auf gewisse Ähnlichkeiten zwischen kompensatorisch gebrauchten artikulatorischen Gesten in mechanischen Perturbationsexperimenten und in „natürlichen“ Sprechgesten hin, und verwenden deshalb Begrifflichkeiten wie „natürliche Perturbation“ für Phänomene wie Koartikulation (wie Geumann, Kroos und Tillmann (1999), hierin Edwards (1985) folgend und sich auf die Wechselwirkung zwischen Kiefer- und Zungenbewegungen beziehend) oder die Veränderung vokaler Intensität (Geumann et al., 1999; Geumann, 2001a). Bei der letzteren „natürlichen Perturbation“ kommt es zu einer Erweiterung des Kieferöffnungsgrades bei Erhöhung der vokalen Intensität, also z. B. bei Rufen (Schulman, 1989; Geumann et al., 1999). Frøkjær-Jensen (1966), Rostolland (1982),Schulman (1985b), Liénard und Di Benedetto (1999), Traunmüller und Eriksson (2000) und Geumann (2001b, 2001a) fanden interessanterweise hierbei eine Kovariation von *f0* und *F1*. Hiervon wird die Korrektheit von Vokalerkennung offenbar nicht negativ beeinträchtigt (Schulman (1985a) und Lindblom und Schulman (1982), letzteres zitiert nach Schulman (1989)), also das Vokalhöhenperzept bleibt trotz *F1*-Änderung erhalten. Dies korrespondiert gut mit den Befunden von Traunmüller (1985), der – unveränderte höhere Formanten vorausgesetzt – für *f0* und *F1*, im gleichen auditorischem Abstand gehalten, aber nach oben verschoben, eine Variation der wahrgenommenen vokalen Intensität ohne Variation der perzipierten Vokalhöhe feststellt.

Trotz des Befundes, dass es ausreicht, *f0* und *F1* anzuheben, ohne die anderen Formanten zu verändern, um das Perzept einer erhöhten vokalen Intensität (oder *vocal efforts*, wie Traunmüller es nennt) hervorzurufen, ist es natürlich so, dass es in „echter“ Sprache nicht ausreicht, diese beiden Parameter zu verändern, um erhöhte vokale Intensität zu erzeugen; stattdessen muss der von den Lippen abgestrahlte Schall tatsächlich eine höhere (akustische) Intensität aufweisen (Black, 1961). Dies wird hauptsächlich durch eine Erhöhung des subglottalen Drucks hervorgerufen (Ladefoged & McKinney, 1963). Eine Erhöhung des subglottalen Drucks bewirkt aber auch eine Erhöhung der Grundfrequenz, da der subglottale Druck Auswirkungen auf die Stimmlippenschwingung hat (siehe Atkinson (1978) für eine abwägende Untersuchung zwischen laryngalen und respiratorischen Einflüssen auf *f0*). Wenn dies der Fall ist, stellt sich die Frage, ob deshalb aus auditiven Gründen (zur Erhaltung der *F1-f0*-Distanz wie z. B. in Traunmüller (1981)) der Öffnungsgrad und damit *F1* variiert werden muss, um die perzipierte Vokalhöhe zu erhalten. Schulman (1989) diskutiert diese Frage ebenfalls, weist aber ebenso darauf hin, dass ein erhöhter subglottaler Druck auch eine Erhöhung des Luftstroms durch die Mundhöhle bedeutet; dieser könnte, da zwischen Luftstrom und Konstriktionsweite ein enger Zusammenhang besteht (Stevens, 1971), für hohe Vokale wie beispielsweise /i,u/, bei denen nicht viel Raum ist zwischen der höchsten Stelle der Zunge und dem Gaumen, bedeuten, dass statt eines Vokals ein eher frikativischer Laut entsteht, so dass für diese Vokale der Öffnungsgrad erhöht werden müsste; um den Vokalraum in der Höhendimension nicht zu verkleinern, müssten dann nachfolgend für alle Vokale der Öffnungsgrad (und damit der erste Formant) erhöht werden. Leider bietet Schulman (1989) keine Zungendaten, so dass er diese Hypothese nicht

selbst untermauern kann[2], aber es ist wahrscheinlich, dass im Gegensatz zu Beißblockexperimenten bei lauter Sprache *nicht* mit der Zunge kompensiert wird, da ja der erste Formant so eindeutig steigt. Aber jedenfalls ist es wahrscheinlich, dass nicht in erster Linie *F1* hier variiert wird, um für die *f0*-Variation mit veränderter vokaler Intensität zu kompensieren und so die Vokalhöhe zu erhalten, sondern dass sowohl die Grundfrequenz als auch der erste Formant durch die notwendige Erhöhung des subglottalen Druckes beeinflusst werden und somit beide mit diesem Faktor variieren. Somit könnte es der Fall sein, dass weniger die Invarianz des *F1-f0*-Abstandes eine Rolle für die korrekte Vokalidentifikation in lauter Sprache spielt, sondern dass Hörer für diese Kovariation in lauter Sprache kompensieren, so wie sie für Koartikulation kompensieren. Dennoch kann nicht ausgeschlossen werden, dass aktiv der *F1-f0*-Abstand versucht wird, zu erhalten; Schulmans Daten (Schulman, 1985a, Seiten 90 bis 91), leider in Hertz angegeben, könnten vermuten lassen, dass der Abstand beider Parameter in Bark sich etwas verringern könnte, denn bei ihm steigt *f0* im Mittel etwa um 100 bis 200 Hz, *F1* jedoch um circa 100 Hz. Auch Traunmüllers eigene Daten zeigen, wenn er *F1* als Funktion von *f0* abbildet, je nach Alters- und Geschlechtsgruppe unterschiedliche Steigungsraten (in Hz) für *F1* (Traunmüller und Eriksson (2000, Seite 3446, Abbildung 6, nicht die in der Einleitung dieser Arbeit gezeigte Abbildung 5) hatte Kinder und Erwachsene beiderlei Geschlechts). Dennoch verlaufen auch diese Daten immer durchaus *in etwa* parallel zu den *f0*-Änderungen. Traunmüller (1990b, Fußnote 3 auf Seite 2018) beschreibt allerdings, dass in der Variation vokaler Intensität innerhalb desselben Sprechers der *F1-f0*-Abstand in Bark *nicht* konstant bleibt.

Das bislang von der „natürlichen Perturbation“ der Variation vokaler Intensität gezeichnete Bild ist eigentlich immer noch nicht vollständig. Wie wollen an dieser Stelle nur sehr kurz auf weitere mögliche Phänomene eingehen. Wie beschrieben, ist es wahrscheinlich, dass sowohl *f0* als auch *F1* mit dem subglottalen Druck variieren. Für die Grundfrequenz ist die bisherige Erklärung somit die, dass sie passiv durch die Erhöhung des subglottalen Druckes beeinflusst ist. Durch Titze und Sundberg (1992) wissen wir aber, dass bei gleichbleibendem subglottalem Druck der von den Lippen abgestrahlte Schalldruckpegel ansteigt, wenn aktiv, also durch die laryngalen Stellkräfte gesteuert, die Grundfrequenz erhöht wird. Es ist also vorstellbar, dass die Erhöhung der Grundfrequenz bei lautem Sprechen nicht nur passiv, sondern auch aktiv vom Sprecher nach oben verschoben wird, um die Effizienz der Schallwandlung zu erhöhen. Ein zweiter Punkt betrifft nun wieder die Rolle des ersten Formanten. Wie Garnier, Wolfe, Henrich und Smith (2008) darstellen, zeichnet sich eine Erhöhung der vokalen Intensität auch dadurch aus, dass die vertikale Kehlkopflage erhöht wird, der Open Quotient sich verringert und die *glottal width* sich erhöht. Wie auch in Kapitel 2.2.6 dargestellt, hat dies Auswirkungen auf die Lage des ersten Formanten. Garnier et al. (2008) testen diese Effekte, indem sie Versuchspersonen die vokale Intensität erhöhen lassen, ohne die Kieferöffnung und die Grundfrequenz zu variieren, und erhalten signifikante Ergebnisse für die erwähnten Maße, also auch für *F1*; abschließend diskutieren sie die Möglichkeit, dass durch die artikulatorischen Änderungen bei erhöhter vokaler Intensität die erste Resonanz des Vokaltrakts an die erste bzw. zweite Harmonische der Grundfre-

[2]Dies gilt auch für Geumann (2001a), die diese Theorie ebenfalls diskutiert.

quenz *getunt* werden könnte, um die Effizienz zu erhöhen (siehe hierzu die Diskussion über klassische Sänger und Formanttuning).

Jedenfalls ist die Variation der vokalen Intensität eine extralinguistische Variationsquelle, die man, wenn man denn möchte, als „natürliche Perturbation“ beschreiben könnte, für die auf die eine oder andere Weise kompensiert werden muss. Insofern ist sie der extralinguistischen Variation der physiologisch bedingten Änderungen der alternden Stimme nicht ganz unähnlich, und insofern ein Analogon.

Eine vom Sprecher nur bedingt beeinflussbare Grundfrequenzveränderung, die durch physiologische Veränderungen hervorgerufen wird, könnte man also ebenso als natürliche Perturbation auffassen. Der Unterschied besteht natürlich darin, dass diese Änderung im Gegensatz zu den bisher genannten „Perturbationen“ nicht plötzlich einsetzt, sondern sich über die Jahre entwickelt. Wie wir in Kapitel 2.2.4 gesehen haben kann man die Grundfrequenzänderungen bei der Königin und auch bei Cooke linear modellieren; wenn wir nur die fallende Grundfrequenz betrachten – eine Eigenschaft, die beiden Sprechern bis zu einem gewissen Alter gemein ist – stellen wir fest, dass die Rate der Änderung vergleichsweise gering ist; wir müssen also davon ausgehen, dass es sich bei diesen Grundfrequenzänderungen um eher subtile Änderungen handelt, so dass z. B. von Jahr zu Jahr betrachtet kaum ein Unterschied feststellen lässt; die Sprecher – so unsere Hypothese denn zutreffen sollte – erfahren also eine „Perturbation“ die nur sehr schwach ausgeprägt und kaum wahrnehmbar sein dürfte. Dennoch – oder gerade deshalb – ist es vorstellbar, dass ein alternder Sprecher für diese „Perturbation“ kompensiert, indem er den Öffnungsgrad in ebenso subtilen Schritten variiert.

Grundgedanke der folgenden akustischen Perturbationsexperimente In diesem Überblick haben wir gezeigt, dass es Befunde aus der Literatur gibt, die vermuten lassen, dass für artikulatorische Gesten im Allgemeinen und für kompensatorische Gesten im Besonderen gilt, dass oft mehrere Artikulatoren im Spiel sind, die sprecherabhängig unterschiedlich eingesetzt werden können, um das gleiche Ziel zu erreichen, so zum Beispiel um unter Perturbation ein zumindest ähnliches auditorisches Feedback zu erreichen wie unter „normalen“ , also unperturbierten Bedingungen. Die im *DIVA*-Modell sogenannten *Goal Regions* auf auditorischer Ebene können vermutlich ebenfalls erreicht werden, indem nicht nur im perturbierten Parameter kompensiert wird, sondern multidimensional. Man muss hier, in Analogie zu den artikulatorischen Befunden, die die Idee der *motor equivalence* stützen, davon ausgehen, dass unterschiedliche Sprecher von den verschiedenen Dimensionen in unterschiedlichem Ausmaß Gebrauch machen werden. Als möglicher Grund für den Gebrauch unterschiedlicher Dimensionen oder Parameter zu Kompensation muss daran gedacht werden, dass möglicherweise die sich widersprechenden Feedbacks aus auditorischer und somatosensorischer Ebene sich so auswirken, dass eine Kompensation in nur einem Parameter möglicherweise nur bis zu einem gewissen Ausmaß „geduldet“ wird, und stattdessen sekundäre Cues – wie die Grundfrequenz für die Vokalhöhe – für das Erreichen einer bestimmten auditorische *Goal Region* mit eingesetzt werden müssen. Eine gewisse Analogie hierzu ist der Gedanke des *feature enhancements*, der u. a. aus den Forschungen

zum Thema vokalinstrinsischer Grundfrequenz stammt. Ähnlich wie dort müssen wir ein gewisses Maß an Unabhängigkeit des sekundären Cues fordern; also eine aktive Kontrolle, um wirklich auf *feature enhancement* schließen zu dürfen.

In der Hauptsache wollen wir als Analogie zu der von uns vermuteten „natürlichen Perturbation" der altersbedingten Grundfrequenzänderungen, die unserer Theorie nach wegen der dadurch bedingten Änderungen in der *Goal Region* für Vokalhöhe langsam zu einer Kompensation und Adaptation bei den Sprechern führt, testen, ob bei künstlichen Grundfrequenzperturbationen in Äußerungen von Wörtern, die in der Vokalhöhendimension lexikalisch konkurrierende Nachbarn haben, bei unvollständiger Kompensation ein Vokalhöhenmismatch auftritt, der sich dahingehend äußern sollte, dass die Sprecher nicht nur mit der *f0*, sondern auch mit dem „klassischen" Vokalhöhenkorrelat *F1* kompensieren. Da die zu erwartenden Effekte vermutlich gering ausfallen werden, sollen zunächst jedoch die Ergebnisse aus Villacorta (2006) und Villacorta et al. (2007) repliziert werden, wo mittels einer *F1*-Perturbation die Vokalhöhe perturbiert wurde, wofür in *F1* unvollständig kompensiert wurde, aber auch unter Einsatz der Grundfrequenz, die eine gegenläufige Bewegung zur *F1*-Bewegung ausführte.

Der Grundgedanke ist also – gegeben, dass alle anderen produzierten akustischen Parameter unverändert bleiben –, dass eine unvollständige Kompensation in *F1* dazu führt, dass das auditorische Feedback nach wie vor *in Richtung* der Perturbation verschoben erscheinen muss. Sollte perzeptuell aber eher ein – wie auch immer gearteter – *F1-f0*-Abstand für die *Goal Region* der Vokalhöhe entscheidend sein, könnten die Sprecher die Grundfrequenz ebenso *in Richtung* der Perturbation verschieben, um den Abstand ähnlich zu halten wie unter der Bedingung ohne Perturbation. Da mit einem Automatismus einer *f0*-Änderung zu rechnen ist, wegen der biomechanischen Abhängigkeiten zwischen Zunge und Kehlkopf, ist hier ein gewisses Ausmaß an unabhängiger *f0*-Kontrolle zu suchen. Analoges gilt für das zweite Experiment, die *f0*-Perturbation: Kompensiert ein Sprecher hierfür unvollständig, erscheint die Grundfrequenz über die auditorische Schleife immer noch *in Richtung* der Perturbation verschoben. Gesetzt, dass diese Grundfrequenzverschiebung einen Einfluss auf die Vokalhöhenperzeption des Sprechers hat, könnte er einen adäquaten *F1-f0*-Abstand dadurch wieder herstellen, dass er *F1* in ebendieselbe Richtung, also *in Richtung* der Perturbation, verschoben produziert.

Dies lässt sich in zwei Hypothesen zusammenfassen:

- Für eine Perturbation in einem der Parameter *f0* oder *F1* kompensieren Sprecher im Mittel unvollständig durch eine Bewegung der Produktion des perturbierten Parameters *in Gegenrichtung* zur Perturbation

- Im Mittel wird der jeweils andere Parameter des Parameterpaares *f0* und *F1* *in Richtung* der Perturbation verschoben produziert. Dies ist – zumindest teilweise – unabhängig von Automatismen.

3.2 Perturbation des ersten Formanten

3.2.1 Methode

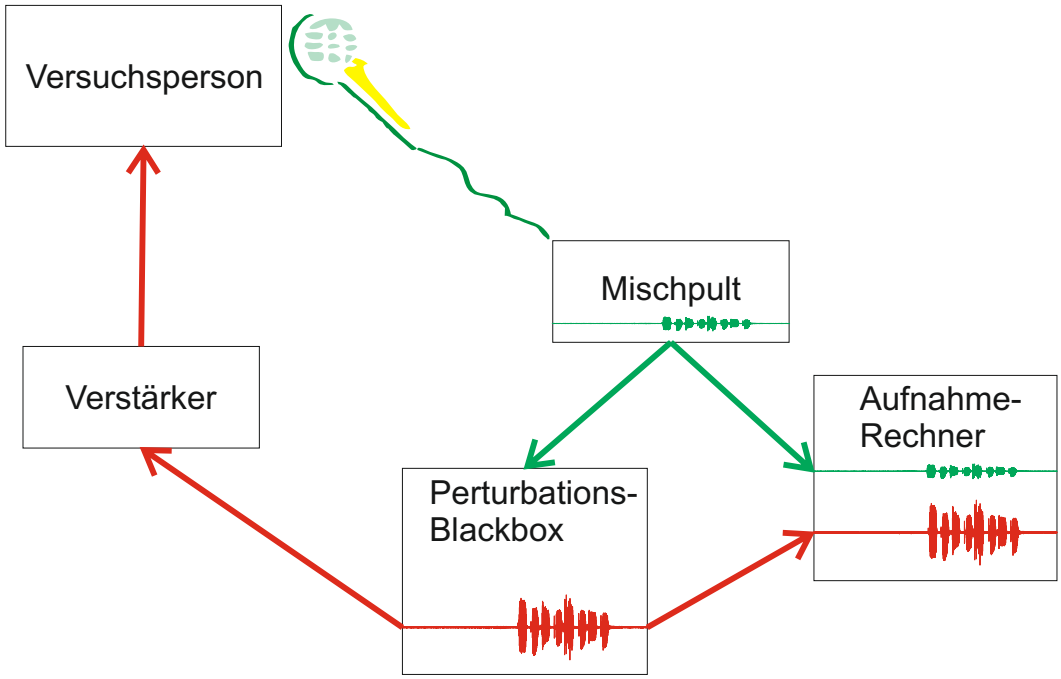

Abbildung 3.2: *Stark vereinfachtes Schaubild des experimentellen Aufbaus für beide Perturbationsexperimente in 3.2 und 3.3. Die Versuchsperson spricht, ihr Signal wird an einem Mischpult aufgespalten. Ein Teil wird aufgenommen, der andere Teil perturbiert und verstärkt an die Versuchsperson zurückgegeben, die sich somit in einer Schleife befindet. Das perturbierte und verstärkte Signal (rot) wird, ebenso wie das (grün dargestellte) Originalsignal aufgezeichnet. Die dargestellten Intensitätsunterschiede spiegeln lediglich die Verstärkung des in der Perturbationsschleife entstehenden Signals wieder; diese Unterscheidung wird auch deshalb bebildert, um klarzustellen, dass die Versuchspersonen in gewisser Weise immer, also auch in den unperturbierten Epochen, einem anderen Feedback als gewöhnlich ausgesetzt sind. Die Latenz, mit der die Versuchsperson beschallt wird, entspricht in etwa der Verschiebung des aus der Perturbationsapparatur kommenden Signals gegenüber dem Originalsignal. Die Analysen in diesem Kapitel beruhen auf die Auswertung der Audioaufnahmen ohne Perturbation (grün).*

Experimenteller Aufbau

Die Aufnahmen fanden in einer schallisolierten Sprecherkabine im Tonstudio des Instituts für Phonetik und Sprachverarbeitung in München statt. Die Versuchspersonen saßen auf einem Stuhl und konnten die Prompts von einem Computerbildschirm, der außerhalb der

schalldichten Kabine direkt an einem Fenster installiert ist, ablesen. Ihr Sprachsignal wurde mittels eines Nackenbügelmikrophons ™Beyerdynamic Opus 54, das sich circa 3 cm links von der midsaggittalen Ebene entfernt befand, aufgenommen und das Signal zu einem Mischpult (™Yamaha O2R) geleitet, wo das Signal aufgeteilt wurde. Einerseits wurde das unveränderte Signal digitalisiert und an die Soundkarte ™M-Audio Delta TDIF in einem Rechner des Typs ™HP Compaq dc7800 CMT PC ALL geleitet, anderseits analog durchgeschleift an die Soundkarte ™M-Audio Delta 44 in einem zweiten Rechner des gleichen Typs, an dem die Formantperturbation stattfand. Nach dieser Formantverschiebung, die im folgenden Unterkapitel beschrieben wird, wurde das perturbierte Sprachsignal wieder an das Mischpult geleitet, dort Digital-Analog-gewandelt und über der Verstärker ™UHER classic Stereo Pre Amplifier UPA-1000 in die Sprecherkabine zu Einsteckohrhörern des Typs ™E-A-RTONE 3A gesendet. Diese Einsteckhörer, die auch für Audiometrieanwendungen benutzt werden, zeichnen sich durch folgende Eigenschaften aus: die Schallwandler, die einen nahezu linearen Frequenzgang erzeugen, befinden sich nicht an den Ohren, sondern hängen an einem Bügel um den Hals der Versuchsperson. Der Schall wird über flexible Schläuche weitergeleitet, an deren anderem Ende durchtunnelte Schaumstoffohrstöpsel aufgesteckt werden können. Diese Ohrstöpsel, vom Hersteller ™E-A-RLink genannt, sind in drei Größen erhältlich und zusätzlich verformbar, so dass sie bestimmungsgemäß jede Gehörgangsgröße akustisch abzukoppeln vermögen. Richtig platziert sind diese Stöpsel, wenn sie verformt so weit als möglich in den Gehörgang eingebracht worden sind, wo man dann den Schaumstoff sich ausdehnen lässt, um die Erfordernisse des festen Halts und der Dämpfung von Außengeräuschen durch das passgenaue Ausfüllen des Gehörgangs zu erfüllen. Laut Herstellerangaben dämpfen die Ohrstöpsel Umgebungsgeräusche um 30-40 dB. Das Signal wurde möglichst laut wiedergegeben, wobei der genaue Wert durch Einstellung während eines auch zum Einpegeln des Aufnahmeequipments benötigten Vortests[3], ermittelt wurde und somit von der Toleranz der Versuchsperson abhängig war.

Formantperturbation mit TransShiftMex

TransShiftMex ist eine von Shanqing Cai, einem Mitarbeiter der *speech communication group* am Massachusetts Institute of Technology (MIT) entwickelte, auf MATLAB basierende Software, die es erlaubt, mit sehr geringer Latenz, also quasi in Echtzeit, formantperturbierte Signale zurückzugeben (Cai et al., 2008, 2010). Hierzu werden mittels einer Schwelle durch eine Kurzzeit-RMS-Analyse stimmhafte von stimmlosen Signalanteilen getrennt und in den stimmhaften Signalanteilen mit *linear predictive coding* (LPC) die Formanten, deren Verlauf online geglättet wird, mit einer LPC-Ordnung von 13 bei männlichen Sprechern und 11 bei Sprecherinnen ermittelt (Xia & Espy-Wilson, 2000). Da hierbei eine Gewichtung der Samples mittels der RMS-Amplitude vorgenommen wird, die Signalanteile aus den Phasen, während denen die Glottis geschlossen ist, bevorzugt, werden Einflüsse subglottaler Resonanzen auf die Formantschätzung verringert. Die eigentlich

[3]Während dieses Vortests musste der /ɑː/-reiche Satz *Barbara saß nah am Abhang, sprach gar sangbar zaghaft langsam, mannhaft kam alsbald am Waldrand Abraham a Sancta Clara* gelesen werden.

Frequenzverschiebung von F1 findet in der komplexen z-Ebene statt, wo durch digitale Filterung Pol-Paare ersetzt werden (Cai et al., 2010).

TransShiftMex erlaubt zeitvariante Formantänderungen sowie Änderungen von Formanten in Abhängigkeit von anderen Formanten, z. B. F1-Perturbationen bei gleichzeitiger, automatischer F2-Änderung . Im vorliegenden Fall jedoch wurden nur statische Verschiebungen um einen bestimmten Wert vorgenommen, und zwar nur in F1.

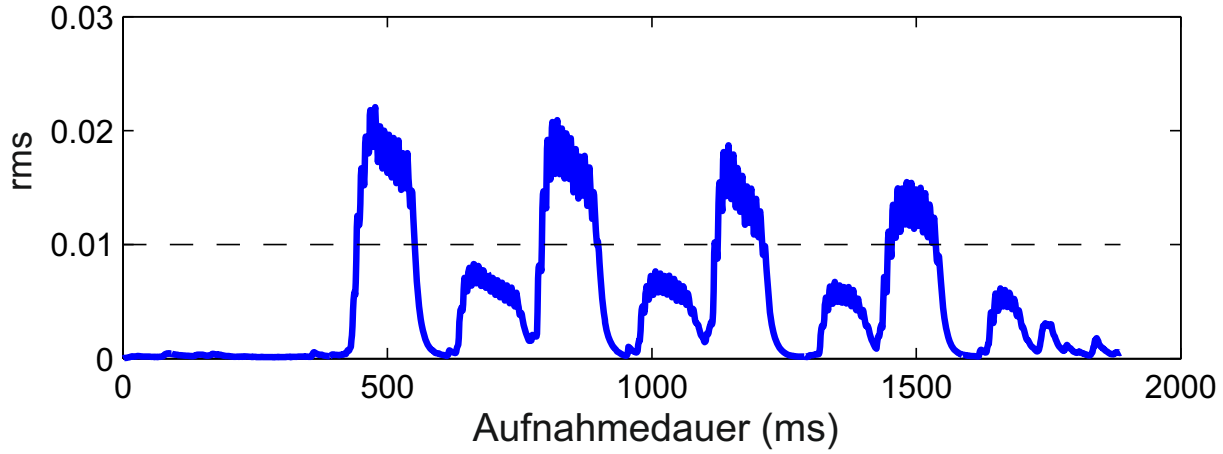

Abbildung 3.3: *Geglättetes RMS-Signal (blau) und RMS-Schwelle (gestrichelte Linie) einer* [beːtnbeːtnbeːtnbeːtn]*-Äußerung des Autors. Die acht Silben, von denen nur die Silben 1,3,5, und 7 die gesetzte Schwelle überschreiten, sind klar zu erkennen.*

Ein besonders kritischer Punkt während des Experimentes ist, dass dieser Formantverschiebungsalgoritmus nur dann applizieren soll, wenn eine bestimmte Intensitätsschwelle überschritten wird, so dass die Formantverschiebungen nur in ausgewählten Signalabschnitten vorgenommen werden. In diesem Experiment sollten beispielsweise nur die Nuclei der ersten Silbe von *beten* in *beten beten beten beten*-Folgen (siehe 3.2.1) betroffen sein. Um dies sicherzustellen, bietet TransShiftMex die Möglichkeit, eine RMS-Schwelle anzupassen; wird diese Schwelle nicht überschritten, appliziert der Formantverschiebungsalgorithmus nicht, so dass im vorliegenden Beispiel Formantverschiebungen eben nur im Nucleus von /beː/-Silben vorkommen, während der Rest des Signals unterhalb der RMS-Schwelle unverändert als Feedback an den Sprecher zurückgegeben werden soll. Die Abbildungen 3.3 und 3.4 zeigen beispielhaft die von TransShiftMex unmittelbar nach jeder Epoche ausgegebenen Abbildungen, die die geglättete RMS-Kurve, die RMS-Schwelle, und die Formantdetektion und -verschiebung wiedergeben.

Die hier abgebildeten Äußerungen zeigen Äußerungen des Autors und dienen als Idealbeispiele. Während des Experiments wurde angestrebt, die abgebildeten Muster zu erhalten, wobei geringe Abweichungen im RMS-Bereich nach oben toleriert wurden, d.h. wenn wenige samples der jeweils zweiten Silbe von /beːtn/ auch über der RMS-Schwelle lagen.

In diesem Fall wurde nur rekalibriert und die Epoche wiederholt, wenn mehr als die (vom Experimentleiter geschätzte) Hälfte der samples des silbischen /n/ über der RMS-Schwelle

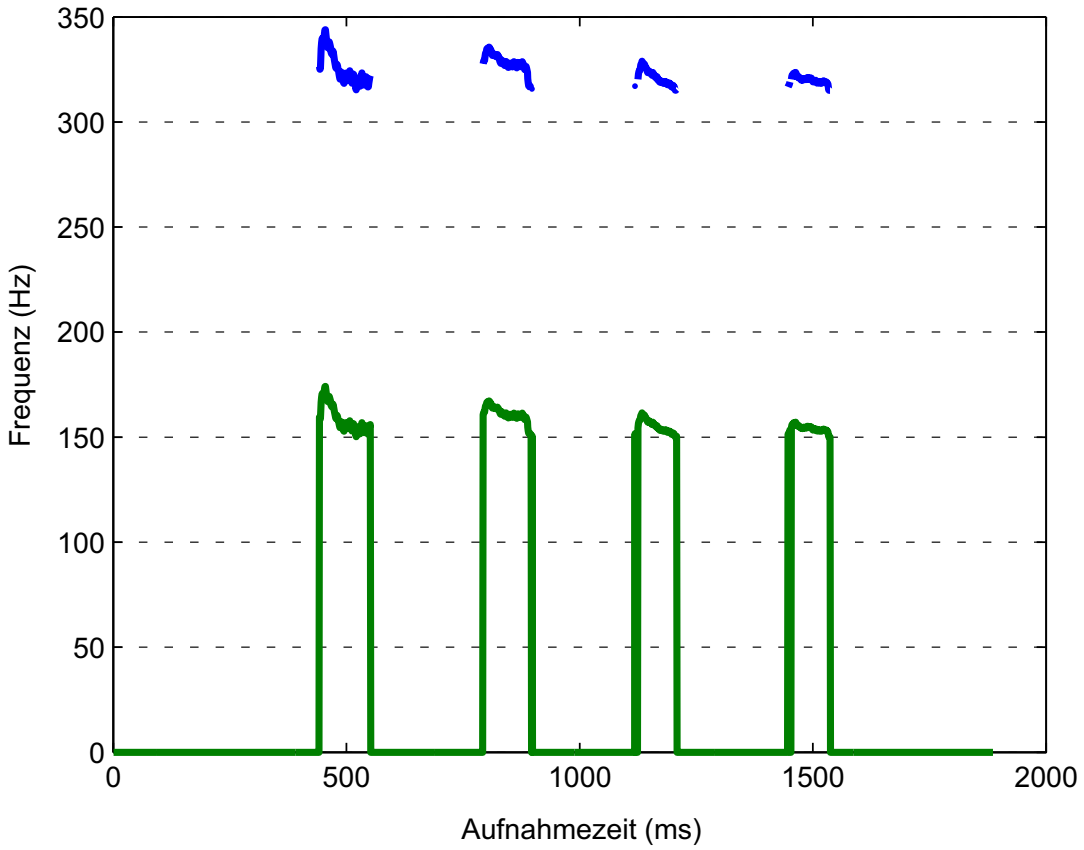

Abbildung 3.4: *Gemessener (blau) und verschobener (grün) erster Formant der gleichen* [beːtnbeːtnbeːtnbeːtn]*-Äußerung des Autors wie in 3.3. Gezeigt wird eine Verschiebung um −200 mel, hier dargestellt in Hz-Skalierung. Die Formant-Detektion und -verschiebung appliziert nur in Signalabschnitten oberhalb einer bestimmten RMS-Schwelle (siehe 3.3). Der Rest des Signals (hier nicht dargestellt) wird unperturbiert in die Feedback-Schleife gegeben.*

lag. Abweichungen der RMS-Werte unter die Schwelle wurden jedoch nicht toleriert, da dies bedeutete, dass keine Formantverschiebung vorgenommen wurde; die (wenigen) betroffenen Epochen wurden konsequent wiederholt. Die Formantverschiebung wurde, wie erwähnt, ohne Abhängigkeit vom zeitlichen Verlauf oder von anderen Formanten durchgeführt. Auch dies zeigen die Abbildungen. Wenn die RMS-Schwelle überschritten wird, wird der erste Formant detektiert und um einen bestimmten Wert verschoben (in 3.4 um −200 mel). Abbildung 3.5 verdeutlicht dies an einem ausgehaltenen [eː], geäußert vom Autor, dessen erster Formant um +200 mel verschoben wurde.

Wie erwähnt wurde das vom Sprecher kommende Signal direkt auf einem zweiten Rechner aufgezeichnet. Auf einem zweiten Kanal wurde zusätzlich das formantverschobene Signal aufgezeichnet, um auch im Nachhinein den Erfolg der Formantverschiebung überprü-

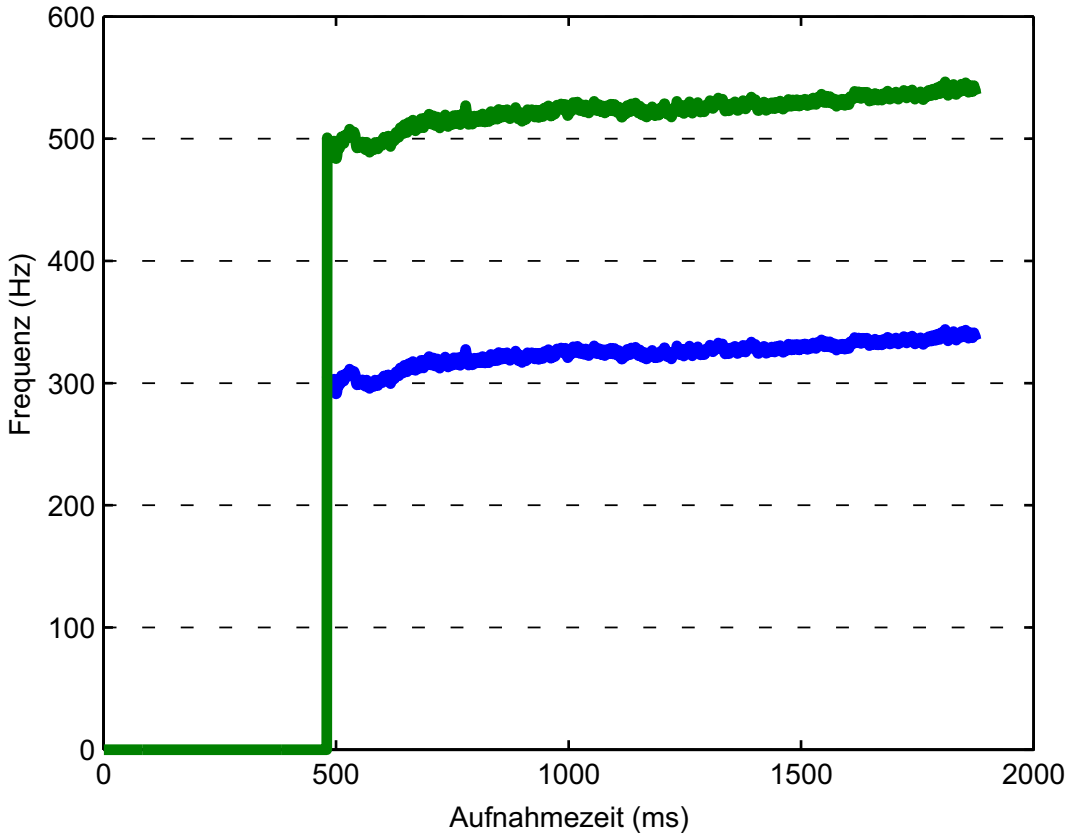

Abbildung 3.5: *Gemessener (blau) und verschobener (grün) erster Formant in einem lange ausgehaltenen* [eː], *gesprochen vom Autor. Der Formantverschiebung betrug +200 mel, hier dargestellt in einer Hz-Skalierung.*

fen zu können. Der Zeitversatz zwischen beiden Signalen ist praktisch identisch mit der Latenzzeit, mit der die Sprecher das Feedback zu hören bekamen. Diese Latenzzeit ist wegen der Zwischenschaltung des Mischpultes etwas höher als vom Entwickler von TransShiftMex angegeben, und beträgt 15 ms, ein Wert, der noch immer weit unter der 30 ms -Schwelle, oberhalb derer Sprecher das Feedback als verzögert wahrnehmen und daraufhin die für *delayed auditory Feedback* typischen Erscheinungen wie Vokaldehnungen, repetitive Konsonantenartikulation, größere Intensität u. s. w. zeigen (Yates, 1963), liegt.

Sprecher

An dem Experiment nahmen 14 Sprecher, davon 5 männlich und 9 weiblich, im Alter zwischen 21 und 43 Jahren teil. Das Durchschnittsalter betrug 28 Jahre. Alle Sprecher wurden am Institut für Phonetik und Sprachverarbeitung rekrutiert. Sieben Sprecher gaben Bayern als Bundesland ihrer Einschulung an (wobei die genaueren Ortsangaben, die ebenfalls

abgefragt wurden, ausschließlich auf eine Herkunft aus dem mittelbairischen Dialektraum schließen lassen), eine Sprecherin Hessen (Südhessen), ein Sprecherin Nordrhein-Westfalen (Bergisches Land), eine Baden-Württemberg (Schwaben), ein Sprecher Schleswig-Holstein und drei Sprecher Sachsen. Bei allen Sprechern wurde vom Experimentleiter ohrenphonetisch überprüft, dass die Sprecher nur gemäßigt von der Standardlautung des Deutschen abwichen, d.h. Sprecher dialekter Aussprachen wurden ausgeschlossen. Dennoch muss die Verwendung von Sprechern, die im bairischen, sächsischen oder schwäbischen Sprachraum aufgewachsen sind, durchaus problematisiert werden, wenn das verwendete Sprachmaterial (siehe 3.2.1) in Betracht gezogen wird, da für alle drei Dialektgebiete nicht auszuschließen ist, dass offenere Realisierungen der Vokalqualität von zugrundeliegend /eː/ tolerierbar sind, da sie im Dialekt auch so produziert werden können (siehe König (1989, Seite 107) für das Schwäbische /ɛː/ für Standard /eː/, Auer, Barden und Großkopf (1993) für die offeneren /eː/-Realisierungen im Sächsischen und König (2009) zum Bairischen, wo z. B. aus /beːtn/ /bɛtn/ werden kann).

Perturbationsprotokoll

Das Experiment bestand aus 125 sogenannten „Epochen“, d.h. 125 mal wurde den Sprechern ein und derselbe Prompt präsentiert, den sie zu lesen gebeten wurden. Dieser Prompt lautete *beten beten beten beten*, und die Sprecher wurden angewiesen, diese Äußerung ohne Satzakzentuierung und möglichst ohne Deklination zu produzieren; diese „monotone“ Äußerungsweise wurde vor dem Experiment kurz eingeübt und sollte mögliche stimmlagenbedingte Variation möglichst ausschließen, wie sie von Liu, Auger und Larson (2010) gefunden worden war: in tieferer Lage der mittleren Grundfrequenz während einer Äußerung war dort auf eine Grundfrequenzperturbation weniger kompensiert worden als in höherer Lage, was hier – soweit möglich – ausgeschlossen werden sollte. Außerdem lautete die Anweisung, die mündlich an die Versuchspersonen weitergegeben wurde, um ihnen auch vorsprechen zu können, bei jedem *beten*-token den Schwa der zweiten Silbe zu elidieren. Ein Grund für die „monotone“ Sprechweise und die Schwa-Elision war, das im vorigen Unterkapitel angesprochene RMS-Schwellen-Kriterium leichter erfüllen zu können; außerdem sollte durch die „monotone“ Intonation eine intonatorisch bedingte *f0*- und auch Formant-Variation möglichst minimiert werden. Die 125 Epochen bzw. 500 (also 125∗4) beːtn-Äußerungen waren in 6 Perturbationsstufen eingeteilt. In den Epochen 1-5 , der sogenannten *Baseline*-Phase, wurde nicht perturbiert, in den Epochen 6 bis 30 wurde abrupt, also ohne Rampenphase, um 100 mel, in den Epochen 31 bis 55 um 200 mel perturbiert, wobei bei der Hälfte der Versuchspersonen der erste Formant nach höheren Werten hin verschoben wurde (diese Versuchspersonen sollen künftig mit *plusminus* benannt werden), bei den anderen 7 Versuchspersonen (künftig: *minusplus*) nach unten. Dieses Vorgehen sollte eventuelle Reihenfolge-Effekte aufdecken. Hierauf folgte in den Epochen 56 bis 75 die sogenannte Rückkehr-Phase, in der nicht perturbiert wurde. Diese Phase sollte durch abruptes Ausschalten der Perturbation dazu dienen, eventuelle Nachfolge-Effekte aufzuspüren und dem Sprecher die Gelegenheit geben, sich allmählich wieder den Werten in der *Baseline*-Phase anzunähern. Daraufhin folgten in den Epochen 76 bis 100 und 101 bis

125 Perturbationen um 100 bzw. 200 mel, diesmal für jeden Sprecher in Gegenrichtung zu den vorangegangenen Formantverschiebungen. Dieses Perturbationsprotokoll ist stilisiert in Abbildung 3.6 dargestellt.

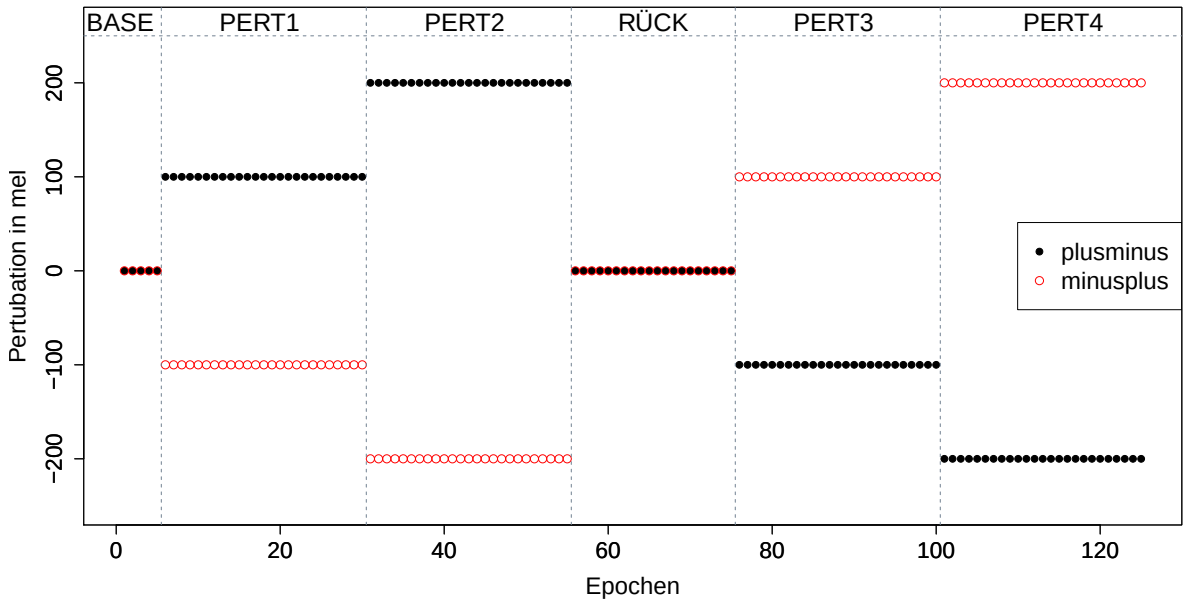

Abbildung 3.6: *Stilisierte Darstellung der Perturbationsphasen (BASE=Baseline-Phase, PERT1-4=Perturbationsphasen, RÜCK=Rückkehrphase) und der in diesen Phasen angewendeten F1-Verschiebungen zwischen −200 mel und +200 mel für die minusplus- (rote, leere Kreise) und die plusminus-Sprechergruppe (schwarze, gefüllte Kreise). Ein Punkt steht für eine Epoche.*

Auf die Rampenphasen wurde aus zwei Gründen verzichtet: erstens sollte das Experiment, das für jede Versuchsperson Perturbationen in zwei Perturbationsstärken (100 bzw. 200 mel) und in zwei Perturbationsrichtungen (*plus* und *minus*) vorsah, nicht unnötig lange werden und somit Ermüdungseffekte nicht zu stark Auswirkungen zeigen lassen; zweitens waren für das zweite Perturbationsexperiment, das als Grundfrequenzperturbationsexperiment ausgelegt war, aus technischen Gründen keine Rampenphasen realisierbar. Aus Vergleichbarkeitsgründen sollten also für das F1-Perturbationsexperiment auch keine Rampen im Versuchsprotokoll eingeplant werden.

Wie erwähnt, wurde der erste Formant bei allen Versuchspersonen um absolute Werte (−200 mel, −100 mel, +100 mel, +200 mel) verschoben. Nun ist es natürlich so, dass ein absoluter Wert an *F1*-Verschiebung für zwei unterschiedliche Sprecher (insbesondere, wenn sie unterschiedlichen Geschlechts sind) nicht den gleichen Effekt auf die Vokalqualität haben, auch wenn dieser Effekt durch Verwendung der mel- statt der Hz-Skala etwas abgemildert wird. Dennoch wurde entschieden, dass diese Vorgehensweise die für dieses Experiment geeignetste ist, da das Hauptziel ist, den Effekt von *F1*-Perturbationen auf die Grundfrequenz

festzustellen. Es ist hierzu nicht notwendigerweise nötig, eine bestimmte Vokalqualität als Ziel der Perturbation anzustreben, zumal es ohnehin zweifelhaft erscheint, ob ein auditorisches Feedback mit z. B. [iː]-Qualität tatsächlich als [iː] vom Sprecher wahrgenommen wird, da dem Sprecher ja eben auch somatosensorisches Feedback zur Verfügung steht und es nicht klar ist, in welchem Ausmaß der einzelne Sprecher auditives und somatosensorisches Feedback gewichtet. Ein unbeantwortete Frage ist, ob die Alternative, die der Formantverschiebungsalgorithmus TransShiftMex bietet, nämlich eine relative Verschiebung des ersten Formanten um einen bestimmten Prozentsatz des gemessenen Formantwerts, tatsächlich einer sprechernormalisierten Vokalqualitätsverschiebung wirklich näher kommt als eine Verschiebung um einen absoluten, mel-skalierten Wert.

Um eine Verschiebung des auditiven Signals von einer [eː]- zu einer [iː]- bzw. [ɛː]-Qualität zu gewährleisten, hätte man in Vortests die mittleren Formantlagen dieser Vokale ermitteln müssen und diese Werte als Zielpunkte der Perturbation erreichen müssen (wobei hiervon wohl auch zumindest der zweite Formant betroffen gewesen wäre, wenn auch in geringerem Ausmaß als *F1*). Auf dieses aufwendigere Verfahren konnte m. E. aus den oben genannten Gründen der Unklarheit über die tatsächliche Perzeption des Sprechers verzichtet werden. Es wurde für ausreichend gehalten, wenn die produzierten Vokale *in Richtung* von [iː] bzw. [ɛː] perturbiert wurden.

Datenauswertung

Zum Zwecke der Datenauswertung wurde eine EMU-Datenbank erstellt. Verwendet wurden hierzu die Aufnahmen des unperturbierten Signals, das direkt vom Sprecher kommend in den Aufnahmerechner geleitet worden war, siehe 3.2. Für die Datenbank wurden die Daten zunächst in *MAuS* (dem *M*ünchner *Au*tomatischen *S*egmentationssystem) (Kipp, 1999; Beringer & Schiel, 1999; Schiel, 1999, 2004) automatisch segmentiert und die so generierte Segmentation nach EMU überführt. Die Grundfrequenz wurde mit dem STRAIGHT-Grundfrequenzdetektionsalgorithmus(Kawahara et al., 1999), dessen Fensterverschiebung auf eine Millisekunde festgelegt ist, ermittelt, die Formanten 1-4 wurden mit forest bei einer Fensterlänge von 30 ms und einer Schrittbreite von einer Millisekunde (um eine Parallelität zu den Grundfrequenzmesswerten zu haben) gemessen; die weiteren Einstellungen bei forest variierten je nach Sprecher und wurden, da mehrere Durchgänge von Messungen durchgeführt wurden, zunehmend dahingehend optimiert, dass möglichst wenige Nullstellen bei F1, F2 und F3 in den Vollvokalen, also in den /eː/s, auftraten. Extrahiert wurden anschließend in emu-R die Formant- und Grundfrequenzwerte zu den mittleren 20% der Vollvokale der *beten*-Äußerungen. Dieses Vorgehen, automatische Segmentation und unkorrigierte Formant- und Grundfrequenzwerte zu benutzen, wurde aus Gründen der Ökonomie gewählt, zumal eine händische Nachkorrektur beim ersten Formanten ohnehin zumeist schwierig ist; stichprobenartige Überprüfung zeigte jedoch, dass dieses Verfahren für diese spezielle Aufgabe vollkommend ausreichend war, da so gut wie keine Fehler auftraten, was auch damit zu erklären ist, dass die Aufnahmen selbst überwacht waren, und - als Folge dieser Überwachung - keine Abweichungen zwischen prompts und Äußerungen vorhanden waren.

Um auch die letzten Ausreißer zu eliminieren, wurden Median-Werte benutzt. Die in die Analyse eingegangenen Werte für *F1* und *f0* entsprechen also den jeweiligen Median-Werten dieser Parameter, entnommen den mittleren 20% des Vollvokals der Äußerungen, d.h. es gibt für jeden der 14 Sprecher pro Parameter für jedes *beten* einen Wert, also vier Werte pro Epoche ($4 * 125 = 500$).

3.2.2 Ergebnisse

Zunächst soll hier die deskriptive Statistik wiedergegeben werden. Als erster Schritt wurden die jeweils vier Werte pro Epoche (vier wegen der vier Wiederholungen des Wortes *beten* pro Epoche) und pro Parameter (*F1* und *f0*) gemittelt, um die Datenlage übersichtlicher zu gestalten. Ein zweiter Schritt bestand darin, alle Werte zur *Baseline* zu normalisieren, d.h. der Mittelwert aller *F1*- bzw. *f0*-Werte aus den ersten fünf Epochen - den *Baseline*-Epochen - sollte nach der Normalisierung 1 sein, und alle anderen Werte aus den übrigen Epochen wurden in Relation zu diesem Mittelwert gesetzt (vergleiche Villacorta (2006, Seite 47, Formel 3.4)). Dies geschieht durch die Formel

$$normparam_{1...125} = \frac{param_{1...125}}{\overline{param_{1...5}}} \tag{3.1}$$

wobei $normparam_{1...125}$ der normalisierte Parameterwert (entweder *f0* oder *F1*) der i-ten Epoche (mit i von 1 bis 125 ist, $param_{1...125}$ der Parameterwert an der durch den Index ausgedrückten Stelle, und $\overline{param_{1...5}}$ der Mittelwert aller Parameterwerte aus den Epochen 1 bis 5.

Betrachtet man Abbildung 3.7, so wird man feststellen, dass eigentlich nur Sprecher LABO_m_MINUS_PLUS in etwa so kompensiert, wie man es von einem *F1*-Kompensierer erwarten würde: Wenn ab Epoche 6 nach unten perturbiert wird, kompensiert er mit einer Verschiebung des ersten Formanten nach oben, kehrt in der *RÜCK*-Phase wieder zu tieferen Werten zurück (in einer etwas überschießenden Reaktion offenbar), und wenn ab Epoche 76 nach oben hin perturbiert wird, produziert er tiefe *F1*-Werte. Interessanterweise scheint selbst dieser Sprecher nicht anders zu reagieren, wenn die Perturbationsstärke erhöht wird.

Andere Sprecher scheinen weit aus weniger auf Perturbation zu reagieren, zumindest in den ersten Perturbationsepochen, zeigen dann aber eine Reaktion, wenn die Perturbation wieder ausgeschaltet wird, die stärker wird, wenn die Perturbationsphasen 3 und 4 eintreten (z. B. MAFE_m_PLUS_MINUS, BABA_w_MINUS_PLUS oder THTH_w_MINUS_PLUS). Dies spricht dafür, dass viele Sprecher einige Zeit benötigen, um überhaupt auf die Perturbationen zu reagieren; dies scheint nicht von der Perturbationsrichtung abhängig zu sein.

Einige wenige Sprecher scheinen sogar die Perturbation zu ignorieren (z. B. ANWE_w_-PLUS_MINUS).

Da offenbar also die Reihenfolge insofern eine entscheidende Rolle zu spielen scheint, als manche Versuchspersonen erst sehr spät beginnen, zu kompensieren, und ihre Kompensationen in den ersten Perturbationsphasen unabhängig von der jeweiligen Perturbationsrichtung eher gering ausfällt, wird ein Maß benötigt, dass die absoluten Werte in ein für

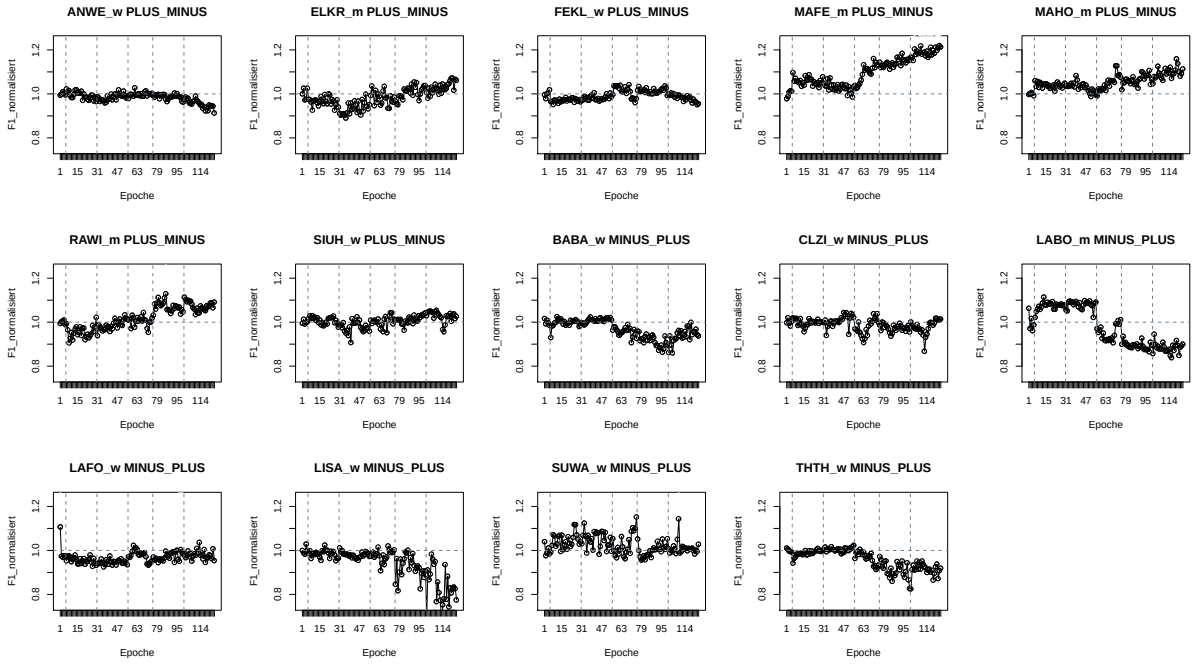

Abbildung 3.7: *Die durchschnittlichen, produzierten F1-Werte der 14 Sprecher in zur Baseline normalisierter Form. Die ersten sieben Versuchspersonen gehören zur PLUS-MINUS-Gruppe, die nächsten sieben zur MINUS-PLUS-Gruppe. Die vertikalen Linien zeigen die Grenzen zwischen den Perturbationsphasen: der BASELINE-Phase folgen die Perturbationsphasen 1 (F1 um 100 mel verschoben) und 2 (F1 um 200 mel verschoben), wobei bei den Versuchspersonen der MINUS-PLUS-Gruppe in Richtung tieferer Werte, bei jenen der PLUS-MINUS-Gruppe nach höheren Werten hin perturbiert wurde. Der daran anschließenden vierten Phase, der RÜCK-Phase, folgen die Perturbationsphasen 3 (100 mel) und 4 (200 mel), wobei die F1-Verschiebung je nach Gruppe nun in die Gegenrichtung zu den Verschiebungen in den ersten beiden Perturbationsphasen erfolgt. Das Geschlecht ist in den Teilabbildungsüberschriften kodiert, ebenso wie die Versuchspersonenkürzel.*

alle Versuchspersonen vergleichbares Maß überführt. Ein solches Maß hat natürlich und zuallererst die Aufgabe, grundsätzlich zu ermöglichen, die Daten aller Versuchspersonen in eine prüfstatistisch auswertbare Form zu bringen, damit sich die Effekte der Perturbationsrichtung nicht gegenseitig aufheben; dies wäre nämlich der Fall, wenn man die Information der Reihenfolge der Perturbation nicht außer Acht lässt. Natürlich kann man getrennt für die zwei Sprechergruppen *MINUSPLUS* und *PLUSMINUS* auf Basis der normierten *f0*- und *F1*-Werte Statistiken rechnen; die Übersichtlichkeit geht hierbei freilich verloren; solche Statistiken, die die Ergebnisse der im Folgenden genannten Tests bestätigen, sind im Anhang A.2 zu finden.

Um ein solches vereinheitlichendes Maß zu gewinnen, erfolgt als nächster Schritt die Übernahme des sogenannten *Adaptive Response*-Wertes aus Villacorta (2006, Seiten 49 f.,

Formel 3.6):

$$AR = \begin{cases} normparam - 1 \mid Perturbationphase = MINUS \bigvee BASELINE \bigvee RÜCK \\ 1 - normparam \mid Perturbationsphase = PLUS \end{cases} \tag{3.2}$$

wobei *normparam* jenem wert entspricht, der mittels der Formel 3.1 für jede Epoche berechnet wurde; je nachdem, zu welcher Perturbationsphase die betreffende Epoche zählt, wird entweder der Wert 1 abgezogen (dies gilt für die Phasen *BASELINE*, *MINUS* und *RÜCK*, somit ist der neue Mittelwert für *BASELINE* nicht mehr 1, sondern 0), oder der Wert *normparam* wird (in der Phase *PLUS*) von 1 abgezogen. Dadurch wird in den perturbierten Phasen ein Wert erzeugt, der positiv ist, wenn F1 in Gegenrichtung zur Perturbation produziert wurde, und der negativ wird, wenn ein Sprecher F1 in Richtung der Perturbation verschoben produziert hat, also nicht gegensteuert. Dieser *Adaptive Response*-Wert hat also den großen Vorteil, dass beide Sprechergruppen, die *PLUS-MINUS*- und die *MINUS-PLUS*-Gruppe, gemeinsam betrachtet werden können (3.8).

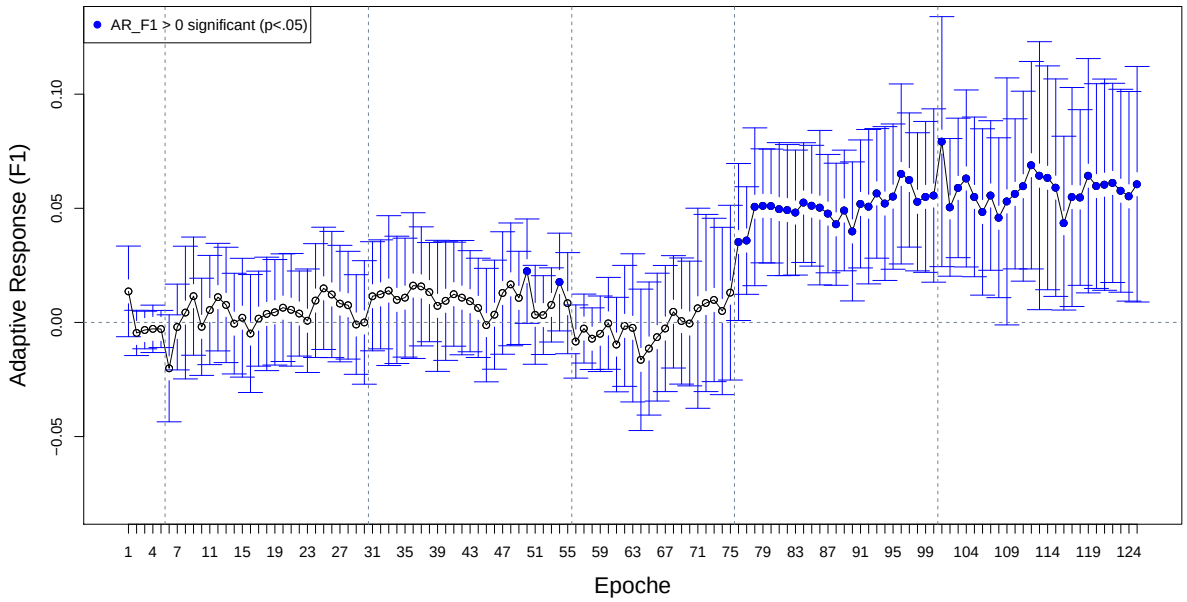

Abbildung 3.8: *Adaptive-Response-Werte für F1 der 14 Sprecher (siehe Formel 3.2). Die Abbildung repräsentiert einen Wert pro Sprecher pro Epoche, wobei die t-Verteilung der Sprecherwerte pro Epoche als blaue Balken gezeigt werden. Der Kreis innerhalb dieser Verteilung entspricht dem arithmetischen Mittel und ist dann blau ausgefüllt, wenn einseitige Einstichproben-t-Tests ergeben haben, dass die Werte für die gegebene Epoche signifikant größer als 0 sind.*

3.8 zeigt, dass über alle Sprecher betrachtet in den ersten Perturbationsphasen (Epo-

chen 6 bis 55) nur wenig kompensiert wird, obschon ein deutlicher Abfall der Werte beim Übergang in die RÜCK-Phase (ab Epoche 56) zu verzeichnen ist. Deutlich ist die recht plötzlich einsetzende Kompensation ab dem Beginn der dritten Perturbationsphase, also Epoche 76 und folgende. Diese Kompensation ist recht deutlich und, über alle Sprecher betrachtet, signifikant, was ein einseitiger Einstichproben-t-test mit $\mu = 0$ ergab, was bedeutet, dass die Werte zum Signifikantsniveau $\alpha = 0.05$ signifikant größer als 0 sind. Dies trifft bei der Perturbationsphase 1 (Epochen 6 bis 30) gar nicht auf, und in Perturbationsphase 2 nur bei den Epochen 50 und 54.

Wie beschrieben, zeigt sich also wenig Bewegung in F1, die als Kompensation gedeutet werden könnte, in den beiden ersten Perturbationsphasen, danach allerdings tiefere *Adaptive Response*-Werte in der Rück-Phase. Aus diesem Grund wurde entschieden, für die Prüfstatistik die F1-Kompensation betreffend die *BASELINE*- und *RÜCK*-Phasen zusammenzulegen (25 Epochen, nämlich Epochen 1 bis 5 und 56 bis 75, im Folgenden *BASIS* genannt) und mit den jeweils 25 Epochen der 4 Perturbationsphasen zu vergleichen. Die Frage, ob die *Adaptive Response*-Werte in den Perturbationsphasen höher seien als in den *BASIS*-Epochen wurde mit Linearen Gemischten Modellen, implementiert im package *lme4* von *R*, entschieden. Die *Adaptive Response*-Werte von *F1* wurden modelliert mit dem Innersubjektfaktor *PERTURBATION* (fünf Stufen: *BASIS, PERTURBATIONSPHASE1, PERTURBATIONSPHASE2, PERTURBATIONSPHASE3, PERTURBATIONSPHASE4*) und *Sprecher* als Random factor[4] . Für alle vier Perturbationsphasen stellt man eine positive Steigung gegenüber der Stufe *BASIS* fest; d.h. für alle Phasen gilt immerhin zumindest eine Tendenz zu höheren Werten, wenn die Sprache perturbiert wurde (*PERTURBATIONSPHASE1*: 0,005; *PERTURBATIONSPHASE2*: 0,012; *PERTURBATIONSPHASE3*: 0,052; *PERTURBATIONSPHASE4*: 0.059). Der Faktor *PERTURBATION* hatte insgesamt einen signifikanten Einfluss auf die *Adaptive Response* von F1 ($\chi^2[4] = 437,1; p < 0,001$). Die Unterschiede zwischen einzelnen Phasen wurden post-hoc mit einem Tukey-Test, wie er im *R*-package *multcomp* (Hothorn, Bretz & Westfall, 2008) implementiert ist, geprüft. Von *Adaptive Response*-Werten der Faktorenstufe *BASIS* unterscheiden sich signifikant jene der Stufen *PERTURBATIONSPHASE2* ($z = 3,3; p < 0,01$), *PERTURBATIONSPHASE3* ($z = 14,3; p < 0,001$) und *PERTURBATIONSPHASE4* ($z = 16,8; p < 0,001$), aber nicht jene von *PERTURBATIONSPHASE1* ($z = 1,3; n.s.$).

Es unterscheiden sich auch nicht die Werte im Vergleich *PERTURBATIONSPHASE1-PERTURBATIONSPHASE2* ($z = 2,0; n.s.$) oder *PERTURBATIONSPHASE3-PERTURBATIONSPHASE4* ($z = 2,2; n.s.$), d.h., es gibt keine Unterschiede, die auf die Perturbati-

[4]Es wurden auch Modelle mit mehreren Faktoren durchgeführt. So gab es ein Modell, bei dem *Adaptive Response* von *F1* (ohne Mittelwertbildung innerhalb der Epochen) sowohl mit dem fünfstufigen Faktor *PERTURBATION* als auch mit dem vierstufigen Faktor *POSITION*, also der Position innerhalb der Epoche (erstes bis viertes *beten* durchgeführt wurde, wiederum mit dem Sprecher als Fehlerterm. *POSITION* hatte hierbei signifikanten Einfluss ($\chi^2 = 13,8; p < 0,001$). Ein weiteres Modell untersuchte, ob der Zwischensubjektfaktor *GRUPPE* (mit den Stufen *MINUSPLUS* und *PLUSMINUS*) einen signifikanten Einfluss ausübte. Dies war nicht der Fall. Siehe zu dieser Frage aber auch die Ergebnisse in Anhang A.2, wo - getrennt nach *GRUPPE* - Statistiken nicht anhand der *Adaptive Response*-Werte von *F1* bzw. *f0*, sondern anhand der normalisierten *F1*- und *f0*-Werte berechnet wurden.

onsstärke (100 mel vs. 200 mel) zurückzuführen wären. Es gibt aber Unterschiede bezüglich des zeitlichen Auftretens der Perturbationsphasen im Verlaufe des Experiments, also zwischen jenen Phasen mit gleicher Perturbationsstärke, wie *PERTURBATIONSPHASE1* vs. *PERTURBATIONSPHASE3* ($z = 13,3; p < 0,001$) und *PERTURBATIONSPHASE2* vs. *PERTURBATIONSPHASE4* ($z = 13,5; p < 0,001$).

Es bleibt also festzuhalten, dass offenbar erst nach einiger Zeit damit begonnen wird, für eine *F1*-Perturbation mittels einer Änderung in *F1* zu kompensieren; diesen Effekt stellt man für die zweite Perturbationsphase auch nur dann fest, wenn man sowohl die *BASELINE*- als auch die *RÜCK*-Phase zusammenlegt, was dadurch gerechtfertigt erscheint, als es im Übergang zwischen der zweiten Perturbationsphase und der *RÜCK*-Phase einen deutlichen Unterschied gibt. Erstaunlicherweise setzt sofort beim Übergang von der *RÜCK*-Phase in die dritte Perturbationsphase ein Kompensationseffekt ein.

Die Perturbationsstärke scheint - entgegen der naiven Erwartung - keine nennenswerte Rolle für die Werte des produzierten ersten Formanten zu spielen. Dies muss bedeuten, da die Perturbationsstärke zunimmt, dass die Kompensation der Perturbation geringer wird, je stärker perturbiert wird.

Bislang haben wir die relative Änderung der unter Perturbation produzierten *F1*-Werte gegenüber den *Baseline-F1*-Werten betrachtet. Wir sollten aber auch, um das Ausmaß der Kompensation abschätzen zu können, ermitteln, wieviel der Perturbation kompensiert wird. Hierzu ermitteln wir pro Sprecher und pro Perturbationsphase (PERT1 - PERT4) die mittlere Abweichung der Produktion von der *Baseline*(in mel) und setzen diese Abweichungswerte in Relation zu den Perturbationswerten ($-200mel, -100mel, 100mel, 200mel$).

Wie Tabelle 3.1 zeigt, gibt es erhebliche sprecherbedingte Unterschiede im Ausmaß der Kompensation; so gibt es in einzelnen Perturbationsphasen Sprecher, die bis zu 55.5% kompensieren, aber es gibt auch einige Sprecher, die der Perturbation folgen. Nur zwei Sprecherinnen folgen aber im Durchschnitt über alle Perturbationsphasen hinweg der Perturbation.

Auch zwischen den Perturbationsphasen gibt es Unterschiede. Tatsächlich wird in den ersten beiden Perturbationsphasen im Mittel wenig kompensiert, wobei dies auch Sprechern geschuldet ist, die in den ersten beiden Phasen der Perturbation folgen, in den Phasen 3 und 4 aber durchaus kompensieren. In den letzten beiden Perturbationsphasen wird kompensiert, aber – wie bereits angedeutet – nur bis zu einem gewissen Plateau. Prozentual ist somit in der letzten Perturbationsphase mit Perturbationsstärke 200 mel weniger kompensiert worden als in Perturbationsphase 3 mit der Verschiebung um 100 mel, wofür im Mittel um immerhin ein viertel kompensiert wurde.

Kommen wir nun zu dem Parameter, der in diesem Experiment am meisten interessiert - der Grundfrequenz. Abbildung 3.9 zeigt wiederum die zur BASELINE normalisierten Werte pro Sprecher, berechnet mit der Formel 3.1.

Daraufhin wurde die Formel 3.2 zur Erzeugung eines *Adaptive Response*-Wertes für *f0* angewandt 3.10. Man beachte, dass nun immer negative Werte entstehen, wenn *f0 in Richtung* der *F1*-Perturbation verschoben wird.

Abbildung 3.10 zeigt die *Adaptive Response* für *f0*. Es wird deutlich, dass *f0* erst in der letzten Perturbationsphase (den Epochen 101 bis 125) sich deutlich von den BASIS-

SprecherIn (Gruppe)	PERT1	PERT2	PERT3	PERT4	Durchschnitt
ANWE (plusminus)	2.9	3.5	-6	-10.5	-2.5
BABA (minusplus)	-1.7	2.4	44.9	16.1	15.4
CLZI (minusplus)	-0.4	1.5	12.2	5.1	4.6
ELKR (plusminus)	14.4	13.8	3	6.1	9.3
FEKL (plusminus)	17	5.1	7.8	-5.4	6.1
LABO (minusplus)	33.4	17.8	47.6	26.1	31.2
LAFO (minusplus)	-25.5	-15.2	15.8	5.8	-4.8
LISA (minusplus)	-8.7	-5.4	41.2	44.9	18
MAFE (plusminus)	-22.7	-5.8	55.5	37.1	16
MAHO (plusminus)	-14.6	-5.7	23.7	17.8	5.3
RAWI (plusminus)	19.2	2.7	30.1	15.8	16.9
SIUH (plusminus)	-4.3	6	6.1	6.9	3.7
SUWA (minusplus)	20.8	11.5	-0.1	-2.6	7.4
THTH (minusplus)	-7.6	1.4	54.3	26.7	18.7
Alle	1.6	2.4	24	13.6	10.4

Tabelle 3.1: *F1-Kompensation in Prozent pro Perturbationsphase und pro Sprecher.*

Werten unterscheidet. Dies bestätigt auch die Prüfstatistik, wiederum in Gestalt eines Linearen Gemischten Modells, diesmal mit den *Adaptive Response*-Werten für f0 als abhängiger Variable und *PERTURBATION* (Stufen: vier Perturbationsphasen sowie *BASIS*, also der Zusammenfassung von *BASELINE* und *RÜCK*) als unabhängiger Variable unter Ausklammerung der sprecherbedingten Unterschiede [5] Auch hier kann festgestellt wer-

[5]Auch *Adaptive Response* von *f0* wurde mit mehreren Faktoren modelliert, siehe die Fußnote auf Seite 109. Für die *POSITION* innerhalb der Epoche wurde ein signifikanter Einfluss auf *Adaptive Response* von *f0* (ohne Mittelwertbildung innerhalb der Epochen) festgestellt ($\chi^2 = 27,5; p < 0,001$), für die *GRUPPE* (*MINUS-PLUS* vs. *PLUS-MINUS*) ebenso ($\chi^2 = 5,5; p < 0,05$), im Gegensatz zu deren Einfluss auf *Adaptive Response* von *F1*. Wie die Abbildungen A.1 und A.2 im Anhang zeigen, nutzen die Sprecher der *MINUS-PLUS*-Gruppe *f0* viel stärker als jene der *PLUS-MINUS*-Gruppe. Wie Abbildung A.2 zeigt, ist

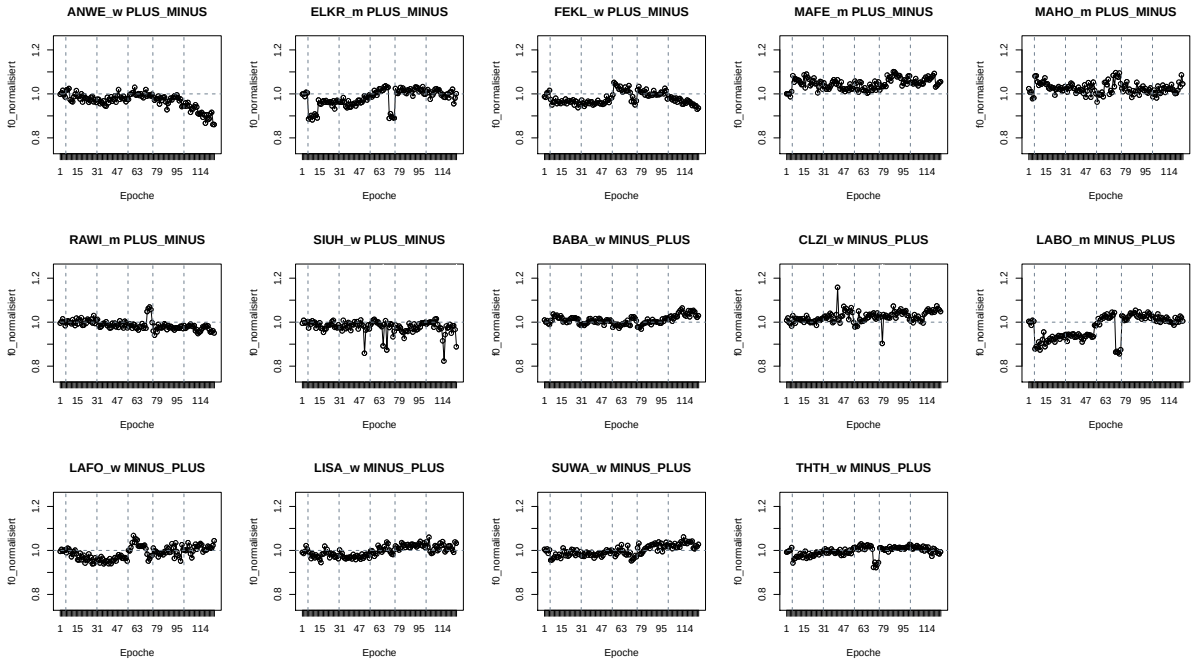

Abbildung 3.9: *Die durchschnittlichen, produzierten f0-Werte der 14 Sprecher in zur BASELINE normalisierter Form. Die ersten sieben Versuchspersonen gehören zur PLUS-MINUS-Gruppe, die nächsten sieben zur MINUS-PLUS-Gruppe. Die vertikalen Linien zeigen die Grenzen zwischen den Perturbationsphasen: der BASELINE-Phase folgen die Perturbationsphasen 1 (F1 um 100 mel verschoben) und 2 (F1 um 200 mel verschoben), wobei bei den Versuchspersonen der MINUS-PLUS-Gruppe in Richtung tieferer Werte, bei jenen der PLUS-MINUS-Gruppe nach höheren Werten hin perturbiert wurde. Der daran anschließenden vierten Phase, der RÜCK-Phase, folgen die Perturbationsphasen 3 (100 mel) und 4 (200 mel), wobei die F1-Verschiebung je nach Gruppe nun in die Gegenrichtung zu den Verschiebungen in den ersten beiden Perturbationsphasen erfolgt. Das Geschlecht ist in den Teilabbildungsüberschriften kodiert, ebenso wie die Versuchspersonenkürzel.*

den, dass in allen vier Perturbationsphasen gegenüber der Basis nierigere Werte vorliegen, da alle vier Steigungen negativ sind (*PERTURBATIONSPHASE1*: -0,01; *PERTURBATIONSPHASE2*: -0,004; *PERTURBATIONSPHASE3*: -0,007; *PERTURBATIONSPHASE4*: -0,018). Insgesamt hat der Faktor *PERTURBATION* einen Haupteffekt auf *Adaptive Response* von *f0* ($\chi^2[4] = 60,9; p < 0,001$), jedoch zeigt der post-hoc Tukey-Test, dass der Effekt (trotz Verwendung der Sufe *BASIS*) bei der letzten Perturbationsphase am stärksten ist ($z = 7,3; p < 0,001$). Auch *PERTURBATIONPHASE1* unterscheidet sich von *BASIS* ($z = 4,0; p < 0,001$); noch knapp signifikant unterschiedliche von der Werten unter der Stufe *BASIS* sind jene der *PERTURBATIONSPHASE3* ($z = 2,7; p < 0,05$), nicht jedoch

der Verlauf der Grundfrequenz bei der *PLUS-MINUS*-Gruppe bestenfalls in der letzten Perturbationsphase wie vorhergesagt *in Richtung* der *F1*-Perturbation.

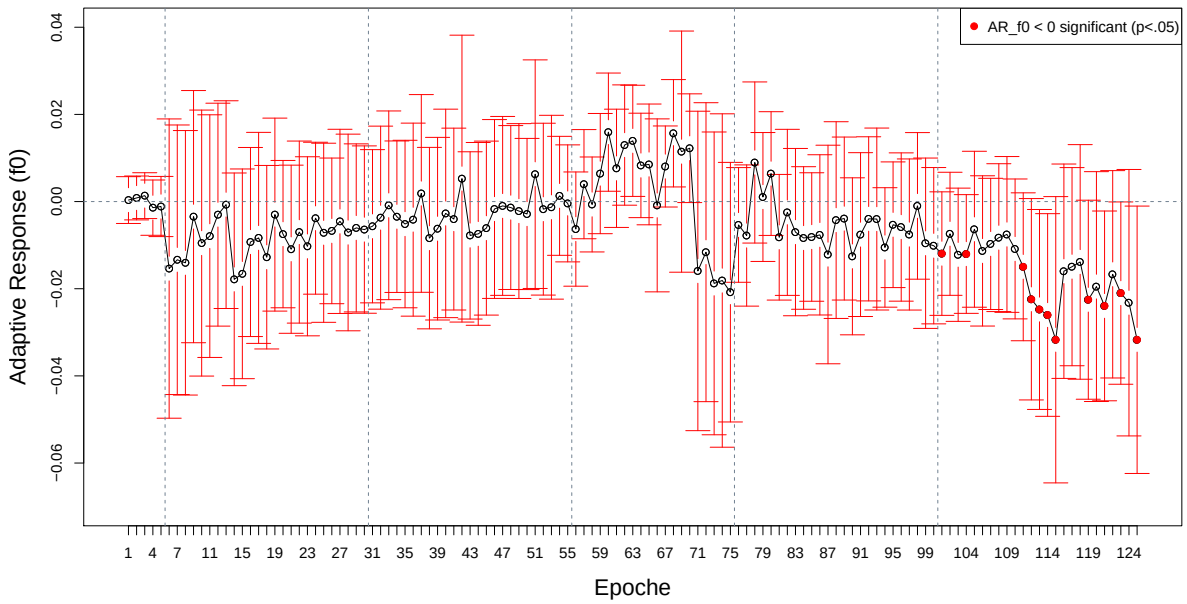

Abbildung 3.10: *Adaptive-Response-Werte für f0 der 14 Sprecher (siehe Formel 3.2). Die Abbildung repräsentiert einen Wert pro Sprecher pro Epoche, wobei die t-Verteilung der Sprecherwerte pro Epoche als rote Balken gezeigt werden. Der Kreis innerhalb dieser Verteilung entspricht dem arithmetischen Mittel und ist dann rot ausgefüllt, wenn einseitige Einstichproben-t-Tests ergeben haben, dass die Werte für die gegebene Epoche signifikant größer als 0 sind.*

jene der *PERTURBATIONSPHASE2* ($z = 1,5; n.s.$).

Besonders interessant ist es, *Adaptive Response* von *F1* und *f0* gemeinsam abzubilden 3.11 und die prüfstatistischen Ergebnisse dabei im Sinn zu behalten. Man erkennt hierbei, dass zwar in den Perturbationsphasen eine Tendenz besteht, dass *f0* dort sinkt und *F1* steigt, aber sich die Parameter möglicherweise ergänzen statt zu kovariieren, denn die Werte für *Adaptive Response* von *f0* sind nicht notwendigerweise dort tiefer, wo jene von *F1* höher sind. So wurde für *PERTURBATIONSPHASE1* berichtet, dass *F1* sich nicht signifikant von den entsprechenden Werten der Stufe *BASIS* unterscheidet, die *Adaptive Response* von *f0* jedoch durchaus. In der zweiten Perturbationsphase ist das Hauptgewicht auf Seiten des ersten Formanten, d.h. die Änderung ist dort ausgeprägt, während *f0* sich nicht (mehr) von den Werten der *BASIS*-Stufe unterscheidet. Hier scheint also, sollte beides als Kompensation für die F1-Perturbation gedeutet werden können, ein entweder-oder vorzuliegen, d.h. die Sprecher ändern zur Kompensation entweder *f0* oder *F1*. In der dritten Perturbationsphase sind beide Parameter signifikant unterschiedlich von den entsprechenden Werten in der *BASIS*-Stufe, und zwar wiederum in gegensätzlicher Richtung, nämlich entgegen der Perturbationsrichtung im Falle des ersten Formanten, und in Richtung der Perturbation im Falle der Grundfrequenz. Wenn nun in der vierten Perturbationsstufe die

Perturbationsstärke ansteigt, wird nicht etwa die Gegenbewegung in *F1* verstärkt (wie die Insignifikanz der *F1*-Werte im Vergleich der Perturbationsphasen 3 und 4 zeigt), sondern die *Adaptive Response* von *f0* verändert sich. Dies sind bereits Hinweise darauf, dass die Hypothese, dass sowohl *F1* als eben auch *f0* zur Kompensation benutzt werden können, stimmen könnte.

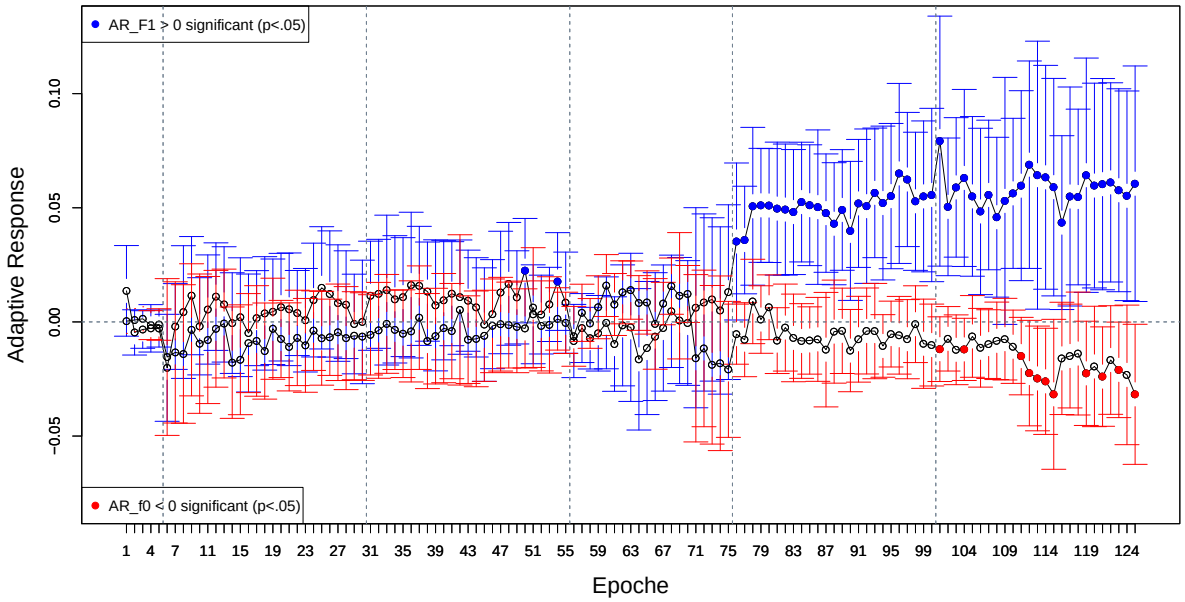

Abbildung 3.11: *Adaptive Response von F1 und f0. Kombination der Abbildungen 3.8 und 3.10*

Villacorta (2006) zeigt mittels einer Abbildung und linearer Modellierung von von *f0* und *F1* abgeleiteter Werte[6], dass in seinen Daten die Tendenz besteht, dass *f0* unter *F1*-Perturbation umso tiefer sinkt (bzw. umso höher steigt), je höher der Sprecher *F1* anhebt (bzw. fallen lässt) (die entsprechende Abbildung Villacortas (Villacorta, 2006, Seite 66) ist im Anhang A.3 zu finden). An dieser Stelle soll zur Anschaulichmachung eine solche Abbildung für die vorliegenden Daten hergestellt und diskutiert werden. Ziel dieses Vorgehens ist es, pro Epoche einen Punkt abzubilden, der die Mittelwerte beider zur BASELINE normalisierter Parameter, *f0* und *F1*, über alle Versuchspersonen bildet und als Koordinaten in einer Ebene bildet. Diese Ebene wird aus den Dimensionen normalisierte *f0* (x) und normalisierter *F1* (y) aufgespannt. Entscheidend wird sein, wie diese Punkte in der

[6]Villacorta (2006) musste, da er einen generellen Anstieg der *f0* in allen Versuchspersonen über den Versuchsverlauf feststellte, d.h. v.a. in den perturbierten Phasen seines Experiments, den über alle Versuchspersonen gemitteltem *f0*-Wert von den Werten seiner Sprecher subtrahieren, um für den generellen Anstieg der Grundfrequenz zu normalisieren (dieses Normalisierungsverfahren wendete er anschließend auch für die normalisierten *F1*-Werte an). In dieser Studie wurde ein solcher *f0*-Anstieg nicht gefunden, d.h. hier wird auf diese Normalisierung verzichtet.

Ebene verteilt sind: Gegeben, dass die *BASELINE* den Ausgangs- und Mittelpunkt der Abbildung bildet (mit den Koordinaten $x = 1.00$ und $y = 1.00$), so sollten die Punkte, die Epochen aus den Perturbationsphase, bei denen *F1* nach unten perturbiert wurde, nach links oben verschoben sein, während jene aus den Perturbationsphasen, bei denen *F1* zu höheren Werten hin perturbiert wurde, nach rechts unten verschoben sein sollten . Da, wie wir festgestellt haben, Änderungen sowohl in *f0* als auch *F1* hauptsächlich in den letzten beiden Perturbationsphasen (den Epochen 76 bis 125) auftreten, wählen wir diese Epochen aus. Dies bedeutet auch, dass für die *PLUS*-Perturbation die Daten der *MINUS-PLUS*-Gruppe, für die *MINUS*-Perturbation jene der *PLUS-MINUS*-Gruppe abgebildet werden.

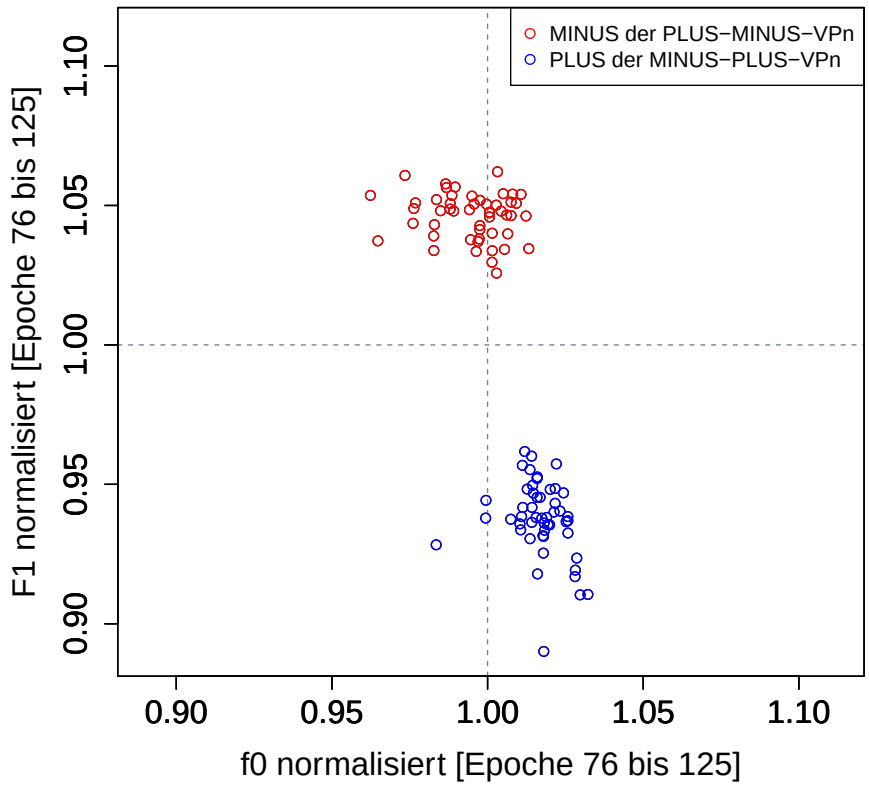

Abbildung 3.12: *F1-Perturbation: Die über die Sprecher der jeweiligen Gruppe gemittelten F1(normalisiert)∼f0(normalisiert)-Werte der Epochen 76 bis 125. Rot: PLUS-MINUS-Versuchspersonen; Blau: MINUS-PLUS-Versuchspersonen.*

Abbildung 3.12 zeigt das erwartete Bild in so weit, als die Punkte der *MINUS-PLUS*-Gruppe (blaue Kreise) tatsächlich rechts unterhalb der Koordinate $x = 1, y = 1$ zu finden sind; lediglich in drei Epochen wird von diesem Muster abgewichen, indem die Grundfrequenz nicht wie vorhergesagt steigt (also in Richtung der *F1*-Perturbation verscho-

ben ist), sondern leicht niedrigere Werte als die der *BASELINE* aufweist. Auch für die *PLUS-MINUS*-Gruppe (rote Kreise) tritt das erwartete Bild insofern zu, als auch dort die Mehrheit der Punkte sich links oberhalb des die *BASELINE* repräsentierenden Punktes $x = 1, y = 1$ zu finden ist. Zwanzig (von fünfzig) Epochen weichen jedoch hiervon ab und weisen, entgegen der Hypothese, höhere *f0*-Werte als zur *BASELINE* auf. In der Tat liegt der Schwerpunkt des Punkteclusters für die *PLUS-MINUS*-Gruppe jedoch nicht nur höher als der der *MINUS-PLUS*-Gruppe, und beide Schwerpunkte dies- und jenseits der *BASELINE* für F1, was für Kompensation für *F1*-Perturbation im *F1*-Bereich spricht, sondern in der Tat auch links verschoben in Relation zu den Werten der *MINUS-PLUS*-Gruppe; die Schwerpunkte beider Gruppen befinden sich links (*PLUS-MINUS*-Gruppe) bzw. rechts (*MINUS-PLUS*-Gruppe) von der *BASELINE* für *f0*, was wiederum ein starker Hinweis darauf ist, dass auch *f0* bei der Kompensation für eine *F1*-Perturbation eine Rolle spielen könnte.

Im Gegensatz zu Villacortas Abbildung A.3 zeigt diese Abbildung, auch weil sie quadratisch, also mit identischen Spannweiten in den Achsen, konzipiert ist, eher eine konzentrische Verteilung der Punkte um einen bestimmten Schwerpunkt, denn eine Verteilung der Werte, die eine lineare Modellierung erlauben würde. Während Villacorta sowohl die Punkte der nach oben perturbierten Sprecher als auch jene der nach unten perturbierten Sprecher linear modellieren kann, und für beide Gruppen einen negativen slope feststellt, was für eine Kovariation beider Parameter spricht, ist dies bei den hier vorliegenden Daten nicht möglich; für beide Gruppen ergibt eine lineare Modellierung nach dem Prinzip der minimierten Quadrate einen R^2-Wert von nahe 0, und in beiden Fällen ist dementsprechend keine Signifikanz dieser Modellierungen feststellbar. Das beide Cluster deutlich unterschieden sind und getrennt voneinander liegen, erscheint auch eine lineare Modellierung über die Daten aller Sprecher, wie Villacorta (2006, Seite 66) sie vornimmt, um die inverse Relation von *F1* und *f0* zu zeigen, nicht sinnvoll, ist aber natürlich dennoch möglich (auf eine Aufnahme der Regressionslinie in die Abbildung wurde aber aus dem genannten Grund dennoch verzichtet). Diese Modellierung ergibt ein $AdjustedR^2 = 0.55$ und eine negative Steigung von -2.73, bei $F[1, 98] = 124, 1; p < 0, 001$.

Diese Betrachtung vergleicht, wie Villacorta, aber zwei unterschiedliche Sprechergruppen zu unterschiedlichen Bedingungen. Wir wollen im Folgenden versuchen, deskriptiv zu untersuchen, wie bei Sprechergruppen in beiden Perturbationsrichtungen reagieren. Hierzu dient Abbildung 3.13, die die normalisierten *f0*- und *F1*-Werte, gemittelt über alle Sprecher einer Gruppe zeigt und diese nach den Stufen *BASIS*, *MINUS* und *PLUS* aufteilt.

Wie man sieht, ist in beiden Gruppen eine Abweichung der *BASIS*, also der Zusammenlegung von *BASELINE* und *RÜCK*, von der *BASELINE* (hier wieder durch den Punkt mit den Koordinaten x=1, y=1 repräsentiert), festzustellen, wobei bei beiden Gruppen *BASIS* (in der Abbildung *B*) dadurch gekennzeichnet ist, dass es von der *BASELINE* für *f0* kaum abweicht, aber deutlich von jener für *F1*; bei der *PLUS-MINUS*-Gruppe, die vor der *RÜCK*-Phase mit einer Perturbation des ersten Formanten nach oben beschallt wurde, liegt *BASIS* deutlich oberhalb der *BASELINE*, bei der *MINUS-PLUS*-Gruppe deutlich unterhalb. Dies ist somit zu begründen, wie bereits in 3.2.2 angedeutet, dass eine erstaunlich große Anzahl an Versuchspersonen auf die ersten beiden Perturbationsphasen so gut wie

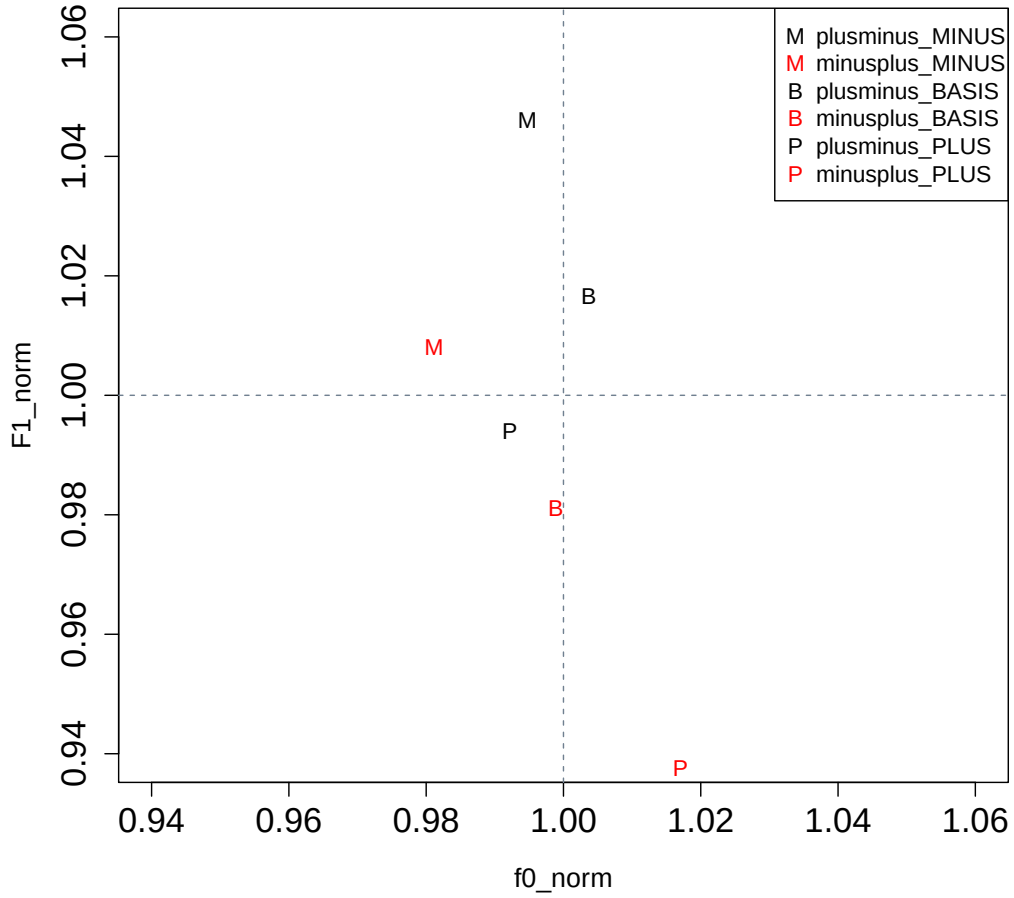

Abbildung 3.13: *F1-Perturbation: Die Mittelwerte über alle Gruppenmitglieder der normalisierten* f0- *und* F1-*Werte für die Stufen (B)* BASIS *(=Kombination der* BASELINE- *und der* RÜCK-*Phase), der (M)* MINUS- *und der (P)* PLUS-*Phase der Sprechergruppen* PLUS-MINUS *(schwarz) und* MINUS-PLUS *(rot). Definitionsgemäß ist die Koordinate x=1, y=1 die* BASELINE.

nicht reagiert (im Sinne einer Kompensation in Gegenrichtung zur Perturbation), jedoch in der Rückphase in eine Lage fällt bzw. steigt, die entfernt von jener der *BASELINE* liegt. Diese Verschiebung erfolgt in Richtung der gerade im Versuchsverlauf zu Ende gegangenen *F1*-Perturbation.

Für die Perturbationsphasen *PLUS* (in der Abbildung *P*) und *MINUS* (hier *M*), über die für die Abbildung unabhängig von der Perturbations*stärke* gemittelt wurde, zeigt sich bezüglich des ersten Formanten durchaus das erwartete Bild: *PLUS* ist gegenüber der *BASIS* nach unten, *MINUS* nach oben verschoben. Auch gegenüber der *BASELINE* trifft dies zu, auch wenn, wie bereits erwähnt, auch hier festzustellen ist, dass die jeweils ersten

zwei Perturbationsphasen (also *PLUS* bei der *PLUS-MINUS*-Gruppe und *MINUS* bei der *MINUS-PLUS*-Gruppe) sehr nahe an der *BASELINE* von *F1* liegen. Dafür sind bei beiden Gruppen entlang der *F1*-Dimension starke Unterschiede festzustellen, wenn man die dritte und vierte Perturbationsstufe (*MINUS* bei der *PLUS-MINUS*- und *PLUS* bei der *MINUS-PLUS*-Gruppe) betrachtet.

Die *MINUS-PLUS*-Gruppe zeigt bezüglich der *f0* das erwartete Bild; *f0* ist sowohl bezüglich der *BASIS* als auch der *BASELINE* immer in Richtung der *F1*-Perturbation verschoben, weist also höhere Werte für *PLUS* und niedrigere Werte für *MINUS* auf. Wählt man *BASIS* (in der Abbildung *B*) als Bezugspunkt, stellt man sogar fest, dass alle drei Punkte (*P*(*PLUS*), *B* (*BASIS*) und *M* (*MINUS*)) fast auf einer Gerade zu liegen kommen, wobei die Abstände zwischen *P* und *B* und zwischen *M* und *B* auch nicht unähnlich sind, also (bei Wahl der *BASIS* als Bezugspunkt) beide Perturbationsrichtungen in dieser Sprechergruppe ähnliche Auswirkungen auf Art und Ausmaß der Kompensation haben.

Anders bei der Sprechergruppe *PLUS-MINUS*: Die dritte und vierte Perturbationsphase (hier also Punkt *M*) führt zu der erwarteten Verschiebung gegenüber *BASIS* nach oben links, was also bedeutet, dass *F1* angehoben und *f0* gesenkt wird. In den ersten zwei Perturbationsphasen ist jedoch nicht nur die Auswirkung auf *F1* im Vergleich zur *BASELINE* wie bereits beschrieben gering, sondern *f0* wird auch *entgegen* der Perturbationsrichtung verschoben, und folgt somit nicht der Hypothese.

Wir stellten bereits durch ein Lineares Gemischtes Modell in der statistischen Analyse anhand der *Adaptive Response*-Werte (siehe Fußnote 5 auf Seite 111) fest, dass es Unterschiede zwischen den Sprechergruppen im Gebrauch der Grundfrequenz zu geben schien. Wir wollen daher im folgenden eine weitere Reduktion der Komplexität vornehmen und uns nur auf den Vergleich der *BASIS*-Stufe mit den vier Perturbationsstufen konzentrieren, dafür aber die Ergebnisse pro Sprecher betrachten. Die Fragestellung hierbei ist, ob es der Fall ist, dass bei der Mehrheit der Sprecher der durchschnittliche *Adaptive Response*-Wert für F1 steigt, während gleichzeitig der *Adaptive Response*-Wert von f0 sinkt.

Die linke Seite der Abbildung 3.14 zeigt pro Sprecher ein Punktpaar, nämlich einen unausgefüllten Kreis für die *BASIS*-Werte und einen ausgefüllten Kreis für Werte aus perturbierten Phasen, wobei diese gemittelt wurden; verwendet wurden in diesem Fall die *Adaptive Response*-Werte für *f0* und *F1*, um über die Perturbationsphasen überhaupt sinnvoll mitteln zu können. Die Paare pro Person sind jeweils mit einer Linie verbunden, deren Steigung in die statistische Auswertung und die Boxplots auf der rechten Seite einflossen. Bei einer knappen Mehrheit der Sprecher ist die Steigung negativ (8 gegenüber 6, siehe auch 3.2); wie im boxplot zu sehen ist, ist eine dieser negativen Werte weit von der Verteilung der anderen Werte entfernt, also ein Ausreißer; zum Zwecke der statistischen Auswertung mittels eines einseitigen t-Tests musste dieser Wert ausgeschlossen werden, um Normalverteilung, eine Bedingung für die Durchführung des t-Tests, zu gewährleisten. Die Normalverteilung der verbleibenden 13 Steigungswerte wurde mittels eines Shapiro-Wilk-Tests getestet und bestätigt ($W = 0,93; n.s.$). Der nun durchgeführte einseitige Ein-Stichproben-t-Test mit $\mu = 0$, der testen sollte, ob die Werte signifikant < 0 sind, ergab $t[12] = -1,9; p < 0,05$, was die Annahme bestätigte.

Tabelle 3.2 listet alle Sprecher im Kontext der Gruppe, der sie angehören, und ihre

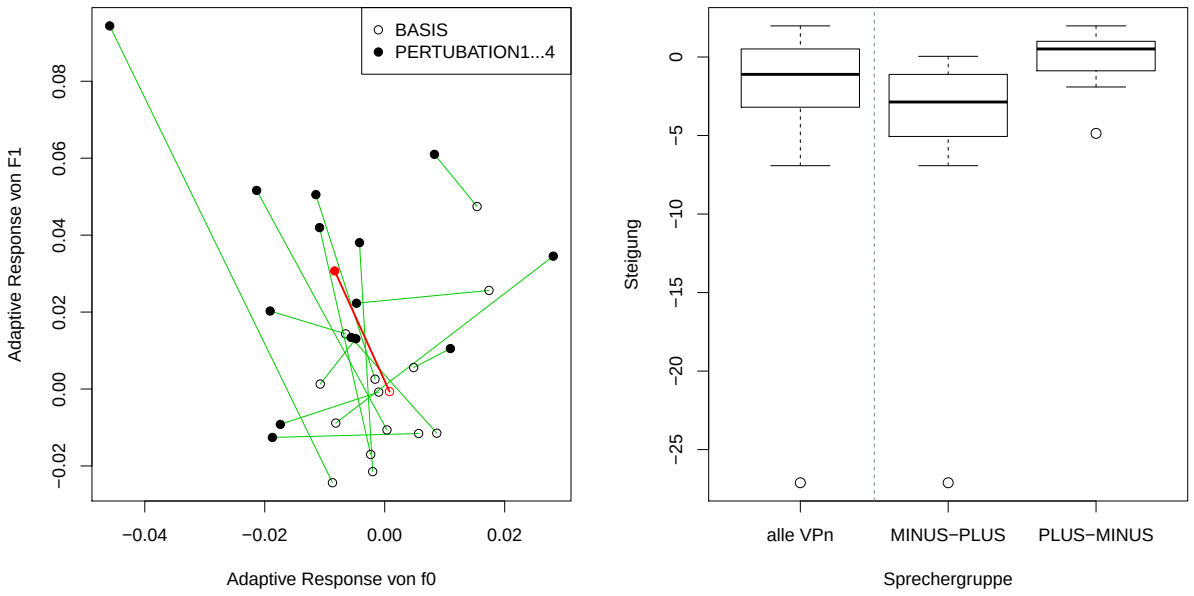

Abbildung 3.14: *F1-Perturbation. Links: Die Mittelwerte über unperturbierte (also* BASIS*-Epochen, unausgefüllte Kreise) und perturbierte Epochen (also der Perturbationsphasen 1 bis 4, gefüllte Kreise) der* Adaptive Resonse*-Werte von* f0*- und* F1 *als Punkte im Adaptive-Response-Raum; je ein Punktepaar pro Versuchsperson. Definitionsgemäß entspricht die Koordinate x=0, y=0 der* BASELINE*. Die Linien verbinden die Werte unperturbierter und perturbierter Epochen. Die rote Linie bildet die gemittelte Steigung aller Versuchspersonen (-3.171) ab. Rechts: Boxplot, also Verteilung, Quantile und Median der Steigungen dieser Linien, für alle 14 Versuchspersonen (linke Seite), und aufgeteilt nach den beiden Sprechergruppen (rechte Seite).*

jeweiligen Steigungswerte (siehe 3.14) auf. Aus der *PLUS-MINUS*-Gruppe weisen 6 von 7 Sprechern eine negative Steigung auf, bei der *MINUS-PLUS*-Gruppe jedoch nur die Minderheit von 2 von 7. V.a. wegen des stark negativen Werts von Sprecher RAWI ist allerding auch in dieser Gruppe der Mittelwert der Steigung negativ (-0.304). Nachdem ein Shapiro-Wilk-Test Normalverteilung für die Daten dieser Gruppe bestätigte ($W = 0,85; n.s.$), wurde ein zweiseitiger Einstichproben t-Test durchgeführt, der ergab, dass die Steigungswerte dieser Gruppe sich nicht signifikant von 0 unterscheiden ($t[6] = -0,34; n.s.$). Für die *PLUS-MINUS*-Gruppe erwies ein Shapiro-Wilk-Test, dass die Daten wegen des starken Ausreißers von Sprecherin BABA nicht normalverteilt sind ($W = 0,66; p < 0,001$). Deshalb und wegen der ohnehin geringer Sprecheranzahl wurde ein nicht-parametrischer Test in Gestalt eines Ein-Stichproben-Wilcoxon-Tests durchgeführt. Dieser ergab, dass die Werte für die Gruppe *PLUS-MINUS* signifikant unterhalb von 0 liegen ($V = 1, p < 0,05$)[7].

[7]Ein mit nach Ausschluss der Sprecherin BABA durchgeführter einseitiger Einstichproben-t-Test ergab

SprecherIn	Steigung	Gruppe	SprecherIn	Steigung	Gruppe
BABA	-27.108	MINUS-PLUS	RAWI	-4.861	PLUS-MINUS
THTH	-6.920	MINUS-PLUS	MAFE	-1.909	PLUS-MINUS
LABO	-3.197	MINUS-PLUS	MAHO	0.151	PLUS-MINUS
LISA	-2.865	MINUS-PLUS	ANWE	0.512	PLUS-MINUS
CLZI	-1.743	MINUS-PLUS	FEKL	0.804	PLUS-MINUS
SUWA	-0.470	MINUS-PLUS	ELKR	1.196	PLUS-MINUS
LAFO	0.042	MINUS-PLUS	SIUH	1.982	PLUS-MINUS

Tabelle 3.2: *Sprecher, Steigung, und Sprechergruppe. Links: die Versuchspersonen der* MINUS-PLUS-*Gruppe. Rechts: die Versuchspersonen der* PLUS-MINUS-*Gruppe.*

Varianzanalysen mit Messwiederholung, durchgeführt für beide Parameter, also *Adaptive Response* von *f0* und von *F1* mit dem *R*-package *ezAnova* (Lawrence, 2011) mit jeweils dem Faktor *PERTURBATION* (mit den Stufen *perturbiert* und *nicht perturbiert*) unter Ausklammerung des durch die Sprecher verursachte Variabilität, ergab einen signifikanten Einfluss im Falle des ersten Formanten ($F[1, 13] = 10,9; p < 0,01$), während bei der Grundfrequenz statistische Signifikanz zum Alpha-Level von 0.05 knapp verfehlt wird ($F[1, 13] = 3,8; p < 0,1, n.s.$) (siehe auch Abbildung 3.15). Man muss hierbei bedenken, dass die Grundfrequenz, wie in 3.2.2 ausführlich besprochen, nur in den späteren Perturbationsphasen eingesetzt wird, hier aber die Daten aller Perturbationsphasen gemittelt wurden.

Als letzte Analyse wollen wir nun noch darstellen, ob es der Fall ist, dass *f0* dann verstärkt eingesetzt wird, wenn der weitere Gebrauch von *F1* geblockt zu sein scheint, wie in Perturbationsphase 4, wo für eine doppelt so starke Perturbation *F1* in etwa genauso verschoben produziert wird wie in Perturbationsphase 3, was ja bedeutet, dass *weniger* in *F1* kompensiert wird. Die Idee dahinter ist die, dass es sein könnte, dass z. B. das somatosensorische Feedback eine weitere Änderung der Kieferöffnung verhindert, und *stattdessen* kompensatorisch *f0* eingesetzt wird. Da die Kompensation in *F1* erst in den letzten beiden Perturbationsphasen konsistent eingesetzt wurde, vergleichen wir hier nur diese beide Phasen miteinander. Abbildung 3.11 scheint zu zeigen, dass die Adaptive Response-Werte von *F1* in beiden Phasen (3 und 4) in etwa gleich bleiben (und die prozentuale Kompensation

- nach Bestätigung der Normalverteilung dieser Daten (Shapiro-Wilk: $W = 0,91; n.s.$) - ein vergleichbares Ergebnis ($t[5] = -2,47; p < 0,05$)

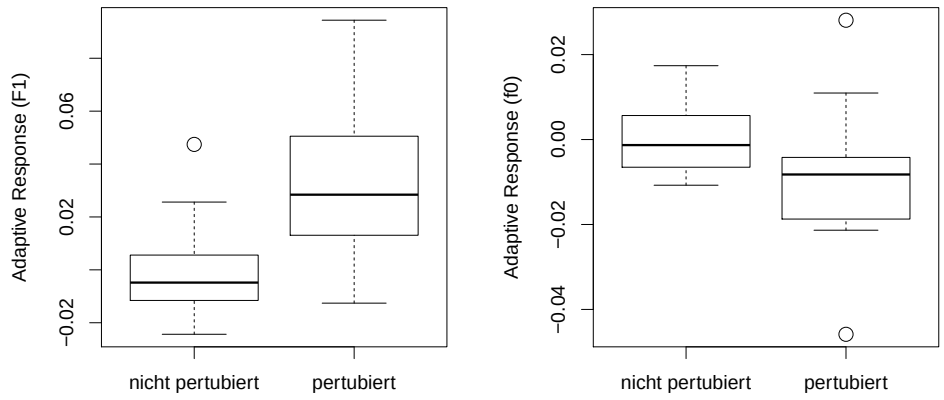

Abbildung 3.15: *F1-Perturbation. Links: Die Verteilung der Mittelwerte über unperturbierte und perturbierte Epochen der* Adaptive Resonse-*Werte von* f0- *und* F1*, berechnet pro Versuchsperson.*

somit sinkt, siehe Tabelle 3.1), während die Adaptive Response-Werte für *f0* in PERT4 stärker abfallen (was bedeutet, dass *f0* stärker genutzt wird) als in PERT3. Um dies zu testen, wurde pro Sprecher die Grundfrequenz in Form der (im Vorzeichen umgedrehten) Adaptive Response sowie die prozentuale Kompensation in *F1* für beide Perturbationsphasen (PERT3 und PERT4) abgebildet und getestet, ob konsistent *f0* dann stärker eingesetzt wird, wenn die *F1*-Kompensation sich verringert.

Wie die Abbildung 3.16 zeigt, gibt es bestenfalls eine kleine Tendenz zu einer Verstärkung des Gebrauchs der Grundfrequenz, wenn der prozentuale kompensatorische Einsatz des ersten Formanten sinkt. Prüfstatistisch wurden die Steigungswerte miteinander verglichen, nachdem Normalverteilung in einem Shapiro-Wilk-Test ($W = 0,92; n.s.$) festgestellt worden war: ein einseitiger t-test, der prüfte, ob die Steigungen unter 0 liegen, ergab keine Signifikanz ($t[13] = 0,53; n.s.$).

Als Alternative wurde ein einfacheres Maß berechnet: für jede Epoche der Perturbationsphasen 3 und 4 wurde die Differenz der Adaptive Response-Werte von *F1* und *f0* berechnet, und diese pro Sprecher und Perturbationsstärke gemittelt, so dass pro Sprecher ein Wertepaar vorlag mit einem Wert für die 100 mel- und die 200 mel-Verschiebung. Ein gepaarter t-Test ergab, dass die Differenzen bei der stärkeren Perturbationsstärke signifikant größer waren ($t[13] = -2,82; p < 0.05$). Betrachtet man wiederum die Differenzen dieser Wertepaare (siehe Abbildung 3.17, rechter Teil), stellt man fest, dass das Ausmaß der Kontrastverstärkung stark zwischen den Sprechern variiert; einige nutzen es offenbar überhaupt nicht.

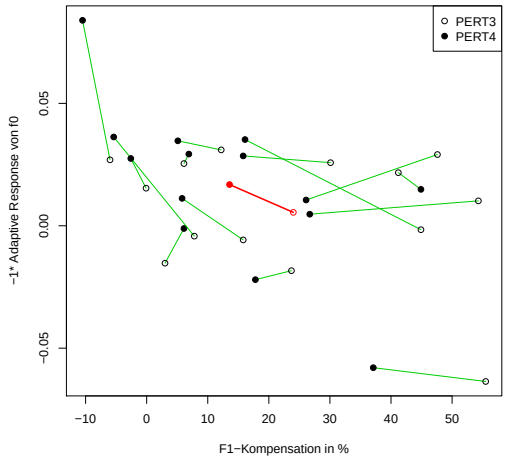

Abbildung 3.16: *F1-Perturbation: Kompensatorischer Gebrauch der Grundfrequenz bei Blockade der* F1*-Kompensation? Auf der x-Achse ist der kompensatorische Gebrauch der Grundfrequenz dimensioniert, während die y-Achse die (im Vorzeichen gedrehte, um der Intuition näherzukommen) Adaptive Response-Werte der Grundfrequenzproduktion zeigt; abgebildet ist ein Punktepaar pro Versuchsperson, das miteinander verbunden ist, wobei die Punkte für die Produktion von* F1 *und* f0 *in den Perturbationsphasen 3 (Perturbationsstärke 100 mel) und Perturbationsphase 4 (Perturbationsstärke 200 mel) stehen. Sollte* f0 *kompensatorisch an Stelle des ersten Formanten eingesetzt werden, müssten die Steigungen zwischen den Datenpunkten für die dritte und vierte Perturbationsphase negativ sein, d.h. wenn prozentual weniger* F1 *genutzt werden kann, sollte der* f0*-Wert steigen. In Rot dargestellt sind die Mittelwerte über alle Versuchspersonen.*

3.2.3 Kurzzusammenfassung der Ergebnisse

Wir wollen hier noch einmal in aller Kürze die wichtigsten Ergebnisse[8] zusammenfassen:

- Die erste Hypothese, dass Sprecher auf eine Perturbation des ersten Formanten mit Veränderungen des ersten Formanten kompensieren, aber unvollständig, konnte bestätigt werden
  - In der ersten beiden Perturbationsphasen ändert sich im Mittel nicht viel an

[8]Im Anhang A.3 ab Seite 200 findet man, zusätzlich zu den hier präsentierten Analysen, eine weitere Analyse der produzierten *F1*- und *f0*-Werte mittels automatischer Klassifikation, die die aus dem *F1*-Perturbationsexperiment stammenden Daten in solche aus perturbierten und unperturbierten Phasen unterscheiden soll; diese Analyse untersucht, inwiefern die hier beschriebenen Mittel, die die Sprecher nutzen, um auf Perturbation zu reagieren, konsistent über Sprecher und Experimentphasen hinweg zu einer über Zufall hinausgehenden korrekten Klassifikation führen, beschränkt sich jedoch auf den Vergleich zwischen den unperturbierten Epochen und den Epochen je einer Perturbationsphase. Da die Ergebnisse dieses Tests ähnliche Resultate wie die hier präsentierten erbrachte, wurde auf eine Darstellung an dieser Stelle verzichtet.

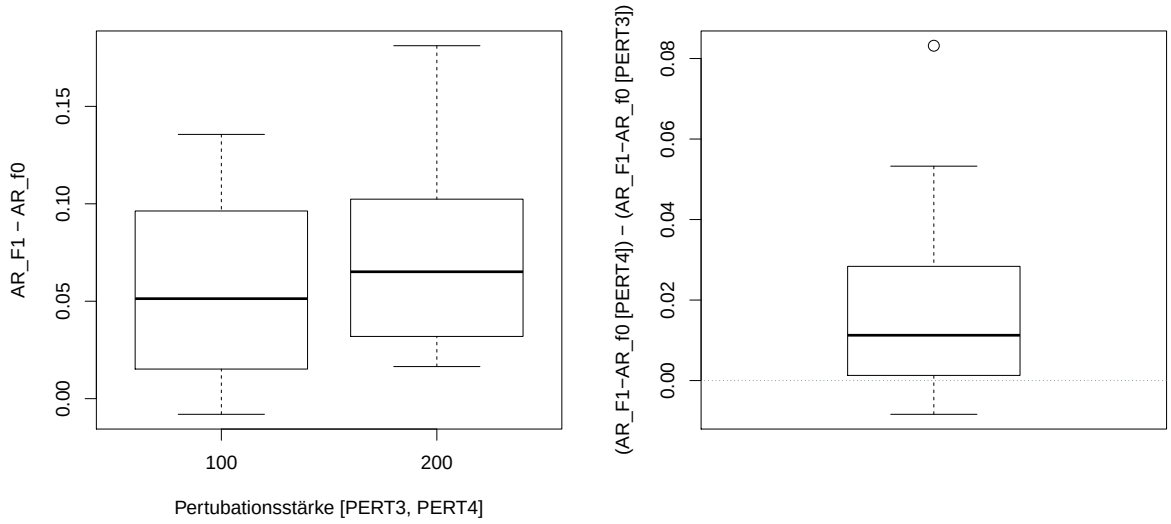

Abbildung 3.17: *F1-Perturbation: Die Mittelwerte der Differenz AR_F1-AR_f0 pro Sprecher und Perturbationsstärke in den Perturbationsepochen 3 und 4; je ein Punktepaar pro Versuchsperson. Rechts: Die Verteilung der Differenzwerte des links gezeigten Maßes, wobei die Werte zur Perturbationsstärke 100 mel von denen der Perturbationsstärke 200 mel subtrahiert wurden; ein Datenpunkt pro Sprecher.*

den produzierten Formantwerten, es wird also scheinbar nicht kompensiert. Es steigt allerdings erheblich die Variabilität der produzierten *F1*-Werte gegenüber derjenigen der *Baseline*-Phase, was darauf zurückzuführen ist, das einige Versuchspersonen durchaus kompensieren, andere aber (hauptsächlich in den ersten beiden Perturbationsphasen) der Perturbation mit ihrer Produktion folgen (die sogenannten *followers*; die meisten davon werden insgesamt im Verlauf des Experiments aber zu Kompensierern), was die Analyse einzelner Sprecher offenbart. Hierbei spielt es allerdings offenbar auch keine entscheidende Rolle, in welche Richtung in diesen ersten Phasen perturbiert wird, da in beiden Sprechergruppen *followers* auftreten, siehe Tabelle 3.1 auf Seite 111.

– Die Rückkehrphase zeichnet sich dadurch aus, dass auf das plötzliche Fehlen einer Perturbation mit einer überschießenden Reaktion geantwortet wird, die produzierten Formantwerte sich also nicht nur von denen der ersten beiden Perturbationsphasen, sondern auch von denen der *Baseline* unterscheiden (siehe A.2 auf Seite 198).

– Ignoriert man die Reihenfolge der Präsentation und kombiniert man die fünf *Baseline*-Epochen mit den 20 *RÜCK*-Epochen zu *BASIS*, ergibt sich auch für die zweite Perturbationsphase ein signifikanter Effekt von Kompensation, nicht aber für die erste Perturbationsphase. Egal, ob man *BASIS* oder *Baseline* als

Grundlage nimmt, unterscheiden sich die Werte des in der dritten und vierten Perturbationsphase produzierten *F1* signifikant; d.h. in diesen beiden Phasen wurde (mit erheblicher Variation zwischen den Sprechern; maximale Kompensation: 55.5%) für die Perturbation in *F1* kompensiert.

- Während die Reihenfolge der Perturbationsphase eine starke Auswirkung auf die produzierten *F1*-Werte hat, ergeben sich keine signifikanten Unterschiede, die auf die Perturbationsstärke zurückzuführen sind; d.h. ungeachtet der Perturbationsstärke wird *F1* in entgegengesetzter Richtung produziert, aber nur bis zu einem bestimmten Wert. Prozentuell nimmt somit die Kompensationsstärke ab, je höher die Perturbationsstärke ist – in den Perturbationsphasen 3 und 4 von 24% auf 13.6%.

- Die Hypothese, dass unter Perturbation des ersten Formanten die Grundfrequenz *in Richtung* der Perturbation verschoben produziert wird – un zwar teilweise unabhängig vom ersten Formanten –, kann bestätigt werden.

  - Für *f0* ergaben sich nennenswerten Effekte durch die Perturbation beim Vergleich zwischen *BASIS* (der Zusammenlegung von *Baseline* und *RÜCK*) und der ersten, der dritten und der vierten Perturbationsphase, wo *f0*, der Hypothese gemäß, *in Richtung* der Perturbation verschoben produziert wurde. Dies trifft aber nicht auf die zweite Perturbationsphase (mit Perturbationsstärke 200 mel) zu. Die Änderungen der Produktion sind wieder deutlicher in den letzten beiden Perturbationsphasen, insbesondere aber in Perturbationsphase 4, siehe beispielsweise Abbildung 3.21.
  - Insbesondere, wenn man die dritte und vierte Perturbationsphase betrachtet, in denen eindeutiger als in der ersten beiden Phasen in *F1* kompensiert wurde, findet man den Effekt, dass *f0* dann tiefer produziert wird, wenn *F1* höher produziert wird, und umgekehrt.
  - Zwischen den Sprechergruppen (PLUSMINUS und MINUSPLUS) gibt es bezüglich des Gebrauchs von *f0* jedoch, im Gegensatz zum Vergleich beider Gruppen zum Gebrauch von *F1*, deutliche Unterschiede. Im Mittel werden in beiden Gruppen die zu tieferen *F1*-Werten perturbierten Äußerungen mit höherem *F1* und niedrigerer *f0* produziert, aber nur die Gruppe MINUSPLUS erhöht ihre *f0*-Werte bei der PLUS-Perturbation, während die PLUSMINUS-Gruppe auch hier tiefere Grundfrequenzen produziert als zu den *BASIS*-Epochen, siehe 3.13. Da die Anzahl der Versuchspersonen pro Sprechergruppe allerdings mit 7 Personen sehr gering ist, kann nicht gesagt werden, ob dieser Effekt ein Gruppeneffekt ist, oder ob er einfach durch individuelles Verhalten einzelner Gruppenmitglieder bedingt ist. Für beide Gruppen gilt jedoch der Reihenfolgeeffekt für *F1*. Während die Daten in der *F1*-Dimension bei der Sprechergruppe PLUSMINUS nach oben verschoben sind, sind sie bei der Gruppe MINUSPLUS nach unten verschoben, d.h. in beiden Gruppen wurde auf die Perturbationsphasen 3 und 4 viel stärker in *F1* kompensiert.

– Wie – neben den beschriebenen Statistiken – Abbildung 3.10 zeigt, scheint es der Fall zu sein, dass *f0* dann verstärkt zum Einsatz kommt, wenn *F1* unverändert bleibt, also wenn möglicherweise eine weitere Kompensation in diesem Parameter nicht mehr möglich ist. Man könnte daraus schließen, dass *f0* die Kompensation in *F1* ergänzt, also nicht einfach damit kovariiert, sondern *stattdessen* genutzt wird, um das Vokalhöhenperzept zu ändern. Dem spricht allerdings schon entgegen, dass in den ersten beiden Perturbationsphasen, wo *F1* in etwa gleich produziert wurde, *f0* unter der schwächeren Perturbation von 100 mel stärker eingesetzt wird. Die Phasen 3 und 4 scheinen aber die Vermutung sich ergänzenden Gebrauchs zu bestätigen, wenn man die mittleren Antworten betrachtet (Seite 113). Eine Analyse des Gebrauchs beider Parameter pro Sprecher in diesen beiden letzten Perturbationsphasen ergab jedoch lediglich eine leichte Tendenz dazu, dass *f0* ergänzend eingesetzt wird. Einige Sprecher tun dies, einige jedoch auch nicht, und statistisch ergibt sich wegen dieser Inkonsistenz keine Signifikanz über alle Sprecher. Es gibt also eine starke Intersprechervariabilität bei der ergänzenden Nutzung beider Parameter - ebenso wie überhaupt bei der Fähigkeit, zu kompensieren. Ein verwandter Test, der ermitteln sollte, ob im Mittel denn die Distanz der Adaptive-Response-Werte von *F1* und *f0* mit zunehmender Perturbationsstärke in den Phasen 3 und 4 anstieg, bestätigte dies: da vorher festgestellt worden war, dass sich die Adaptive Respone von *F1* nicht in den Phasen 3 und 4 signifikant unterscheidet, kann man davon ausgehen, dass doch (zumindest von einigen Sprechern) *f0* unabhängig zur Kontrastverstärkung eingesetzt wird.

Im Folgenden wollen wir untersuchen, ob für Grundfrequenzperturbation auch unvollständig kompensiert wird, und – falls dies der Fall sein sollte – ob dadurch die Vokalhöhenperzeption betroffen wird, was wir dadurch abschätzen wollen, indem wir ermitteln, inwiefern der erste Formant *in Richtung* der Grundfrequenzperturbation verschoben wird oder nicht.

3.3 Perturbation der Grundfrequenz

3.3.1 Methode

Experimenteller Aufbau

Der experimentelle Aufbau glich in wesentlichen Punkten dem Aufbau, der für das F1-Perturbationsexperiment verwendet worden war, d.h. die Aufnahmen fanden in einer schallisolierten Sprecherkabine im Tonstudio des Phonetikinstituts statt, und zwar direkt im Anschluss an das Formantperturbationsexperiment. Die Versuchspersonen saßen wiederum auf einem Stuhl und konnten die Prompts von einem Computerbildschirm, der außerhalb der schalldichten Kabine direkt an einem Fenster installiert ist, ablesen. Wieder wurde das Nackenbügelmikrophon ™Beyerdynamic Opus 54 verwendet, und das Signal wurde zu einem Mischpult (™Yamaha O2R) geleitet, wo das Signal aufgeteilt wurde. Einerseits wurde das unveränderte Signal digitalisiert und an die Soundkarte ™M-Audio Delta TDIF in einem Rechner des Typs ™HP Compaq dc7800 CMT PC ALL geleitet, anderseits analog durchgeschleift zu einem sogenannten Harmonizer des Typs TC-HELICON® VOICELIVE2. Nach der dort stattgefundenen Echtzeit-Grundfrequenzverschiebung wurde das perturbierte Sprachsignal wieder an das Mischpult geleitet, dort Digital-Analog-gewandelt und über einen ™UHER classic Stereo Pre Amplifier UPA-1000 verstärkt in die Sprecherkabine zu Einsteckohrhörern des Typs ™E-A-RTONE 3A gesendet. Das Signal wurde möglichst laut wiedergegeben, wobei der genaue Wert durch Einstellung während eines auch zum erneut notwendig gewordenen Einpegeln des Aufnahmeequipments benötigten Vortests ermittelt wurde und somit auch bei diesem Experiment von der Toleranz der Versuchsperson abhing.

Grundfrequenzperturbation

Der Harmonizer TC-HELICON® VOICELIVE2 dient eigentlich dazu, einer einzelnen Gesangsstimme eine oder mehrere Nebenstimmen hinzuzufügen, und zwar in einem musikalisch begründeten Abstand, um z. B. einen Zwei-, Drei- oder Mehrklang zu erzeugen, also um die Erzeugung von Harmonien im musikalischen Sinne, d.h. von Übereinanderschichtungen mehrerer als *Töne* wahrgenommene Frequenzen, die in diesem Fall Grundfrequenzen von Stimmen sind. Dies ist prinzipiell technisch sehr einfach realisierbar, indem das Spektrum des gesamten Signals um einen bestimmten Faktor angehoben bzw. gesenkt wird, und viele einfache *pitch shifter* arbeiten auf diese Weise. Dies führt - insbesondere bei größeren Tonabständen - zu unnatürlichen Stimmwahrnehmungen, die manchmal als Effekt sogar erwünscht sind (Schlumpf- bzw. Monster-Stimmen). Für viele Gesangsanwendungen, aber eben auch für dieses Experiment ist es jedoch entscheidend, dass nur die Grundfrequenz, nicht aber die sonstigen spektralen Eigenschaften der Stimme verändert werden. Dazu ist eine Trennung von Quell- und Filtersignal notwendig, die das Gerät TC-HELICON® VOICELIVE2 bietet. Leider ist die technische Realisierung des von dem Gerät produzierten *formanterhaltendem pitch-shifting* nicht in der Dokumentation beschrieben, so dass das Gerät in doppeltem Wortsinne eine black box bleiben muss - ein Signal wird ein-, und ein Grundfrequenz-verschobenes Signal wird formanterhaltend wieder ausgegeben. Die Erhal-

tung der Formanten wurde vor Beginn der Experimentreihe durch einige, mit zwei Kanälen, die sowohl das unperturbierte wie das perturbierte Signal erhielten, arbeitenden Aufnahmen überprüft und bestätigt.

Das Gerät ist in der Lage, auch nur *eine* grundfrequenzverschobene „Stimme“ wiedergeben zu können (unter Weglassung der originalen Stimme), so dass das Gerät für die in diesem Experiment angedachte Aufgabe die ausreichende Funktionalität besaß. Leider gibt es jedoch eine wichtige Einschränkung: wegen der erwähnten Ausrichtung des Gerätes auf musikalische Zwecke kann die Grundfrequenzverschiebung nur in Halbtonschritten vorgenommen werden, d.h. die kleinste Perturbationsgröße ist ein Halbton.

Die am Unterschied der beiden aufgezeichneten Signale (direkt und perturbiert) gemessene Latenzzeit ist wegen der Zwischenschaltung des Mischpultes und der dadurch bedingten mehrfachen AD/DA-Wandlung vergleichsweise hoch [9] und beträgt 40 ms, ein Wert, der über der von Yates (1963) genannten 30 ms -Schwelle, liegt. In der Tat zeigten zwei Sprecher gelegentlich die für *delayed auditory Feedback* typischen Erscheinungen, hier in Form von größerer Intensität und gelgentlichem „Versprechen“. Die betroffenen Epochen wurden wiederholt, wobei allerdings vom Versuchsleiter eine relativ große Toleranz für die Akzeptanz der Äußerungen ausgeübt wurde.

Sprecher

Es nahmen die gleichen 14 Sprecher wie im Formantperturbationsexperiment teil, siehe die Beschreibung in 3.2.1.

Perturbationsprotokoll

Der Aufbau des Perturbationsprotokoll war demjenigen der Formantperturbationsexperiments sehr ähnlich. Wiederum wurden 125 Epochen von /be:tnbe:tnbe:tnbe:tn/-Äußerungen aufgezeichnet, wobei ab der sechsten Epoche *f0* perturbiert wurde. Die Epochen 56 bis 75 waren die Rückkehrphase, also ohne Perturbation. Wieder waren die Sprecher in eine *plusminus*- und eine *minusplus*-Gruppe à sieben Personen unterteilt. Das Perturbationsmuster in 3.18 ähnelt stark dem in 3.6.

Wie die Abbildung 3.18 zeigt, wurden in diesem Experiment eine Perturbationsspannweite von ± 2 Halbtönen gewählt.

Ein großer Unterschied zum *F1*-Perturbationsexperiment war jedoch, dass den Versuchspersonen vor jeder Epoche eine Aufnahme der eigenen Stimme vorgespielt wurde, in der eine /be:tnbe:tnbe:tnbe:tn/-Äußerung zu hören war, die unmittelbar vor dem Experiment vom Versuchsleiter ausgesucht worden war. Diese Aufnahmen stammten jeweils aus der Baseline-Phase (Epochen 1-5) des F1-Perturbationsexperiment, und wurden ausgewählt in Hinblick auf die möglichst beste Erfüllung der für dieses Experiment gültigen Kriterien, also danach, ob

- eindeutig [be:tnbe:tnbe:tnbe:tn] produziert wurde

[9] Die Latenzzeiten der unterschiedlichen AD/DA-Wandlungen und des *f0*-Perturbations-Gerätes addieren sich

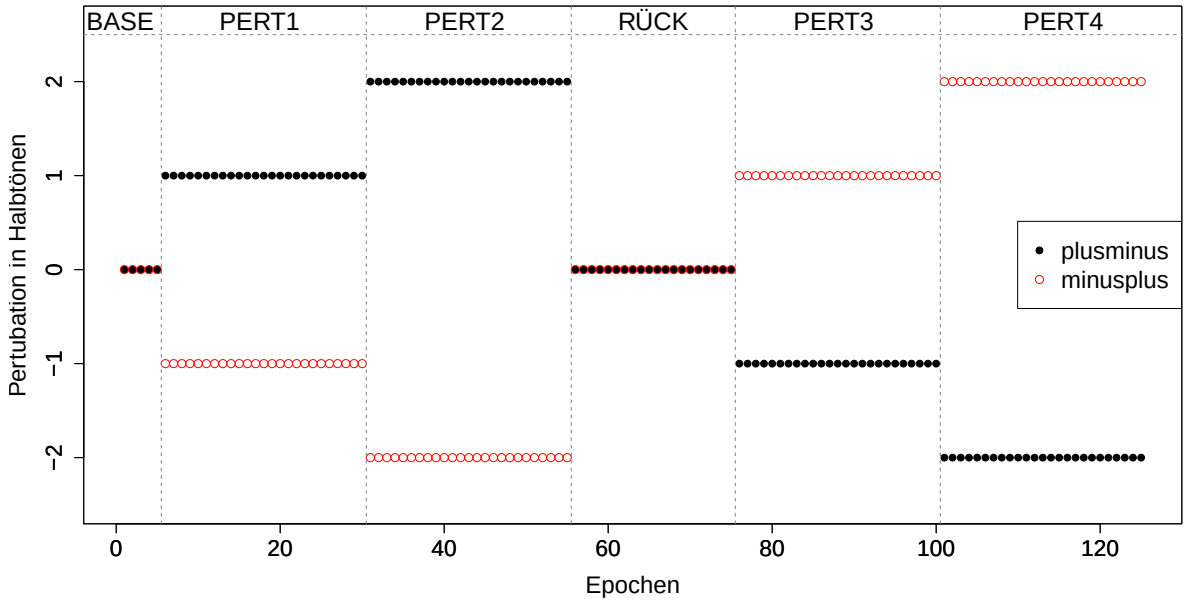

Abbildung 3.18: *Stilisierte Darstellung der Perturbationsphasen (BASE=Baseline-Phase, PERT1-4=Perturbationsphasen, RÜCK=Rückkehrphase) und der in diesen Phasen angewendeten f0-Verschiebungen zwischen −2 Halbtönen und +2 Halbtönen für die minusplus- (rote, leere Kreise) und die plusminus-Sprechergruppe (schwarze, gefüllte Kreise).*

- die Äußerung „monoton“, also ohne Satzakzentuierung und ohne Deklination der Grundfrequenz produziert wurde

Dieses Vorgehen ist vergleichbar mit dem in Liu, Auger und Larson (2010), wo auf diese Weise versucht wurde, einen Referenzwert vorzugeben.

Die gewählte Aufnahme wurde unmittelbar vor jeder Äußerung in den Epochen des *f0*-Perturbationsexperiments vorgespielt, und (die mündlich vom Versuchsleiter an die Versuchspersonen weitergegebene) Aufgabe der Sprecher war, die eigene Äußerung „so gut als möglich nachzuahmen“, wobei darauf verwiesen wurde, dass diese Anordnung sowohl die Identität der Wörter als auch die „Lage der Stimme“ betraf.

Ein weiterer, entscheidender Unterschied zwischen den Experimenten ist sicherlich, dass die Perturbation im *F1*-Experiment nur das /eː/ aus *beten beten beten beten* betraf, im *f0*-Experiment jedoch immer applizierte, also bei jeder Äußerung, die die Versuchsperson machte, was z. B. auch Zwischenfragen usw. betraf. Es wurde durch Befragen der Versuchspersonen nach den Aufnahmen auch deutlich, dass diesen bei dem *f0*-Perturbationsexperiment die Perturbierung ihrer Sprache offenbar wesentlich bewusster war als im *F1*-Perturbationsexperiment.

Datenauswertung

Die Datenauswertung erfolge genauso wie in 3.2.1 beschrieben, d.h. über Medianwerte für *f0* und *F1* der mittleren 20% des Vollvokals der *beten*-Äußerungen, so dass es wieder 500 Werte (= 125 Epochen * 4 *beten*-Tokens) pro Sprecher auszuwerten galt. Für die Bestimmung des Ausmaßes der Kompensierung im *f0*-Bereich wurden die Grundfrequenzdaten teilweise in Halbtöne umgerechnet. Dies geschah mit der Formel

$$12 \times \log_2(\frac{f0}{\overline{f0[BASELINE]}}) \tag{3.3}$$

wobei $f0$ der gegebene Grundfrequenzwert ist, $\overline{f0[BASELINE]}$ der Mittelwert der Grundfrequenzen des selben Sprechers aus den *BASELINE*-Epochen (1 bis 5), und $\log_2$ der Logarithmus zur Basis 2.

3.3.2 Ergebnisse

Zunächst präsentieren wir in Abbildung 3.19 die normalisierte Grundfrequenz, errechnet nach Formel 3.1 (siehe Seite 106). Relativ viele Sprecher, insbesondere aber Sprecher *MAHO* und *THTH*, reagieren fast augenblicklich und vergleichsweise stark kompensatorisch auf die Grundfrequenzperturbation, was aus den treppenförmigen Verläufen bei diesen Sprechern ersichtlich wird. Einige Sprecher reagieren aber auch sehr wenig, einige folgen sogar der Perturbationsrichtung, d.h. sie verändern die Grundfrequenz ihrer Äußerungen in die Richtung, in die auch perturbiert wurde (und verstärken damit die Unterschiede, anstatt sie auszugleichen).

Wegen der Verwendung von Halbtonschritten bei der Perturbation besteht die Möglichkeit, sehr einfach das Ausmaß der Kompensation für diese Perturbation zu quantifizieren, indem Formel 3.3 angewendet wird. Hierfür wurde für jede Perturbationsstufe (−2 Halbtöne,− 1 Halbton, keine Perturbation,+1 Halbton,+2 Halbtöne) pro Sprecher ein Mittelwert der Grundfrequenzveränderung gegenüber der *BASELINE* errechnet und - je nach Kompensationsstärke und -richtung - der prozentuale Anteil an Gegenbewegung zur Perturbation errechnet. Abbildung 3.20 präsentiert die Kompensation über alle Sprecher hinweg, dargestellt als Verteilung der Kompensation in Halbtönen bzw. als prozentualer Anteil. Einige Sprecher kompensieren im Mittel fast vollständig (und überkompensieren offenbar sogar gelegentlich, siehe 3.20, wo bei den Perturbationsstufen −1 und 1 Kompensationen über 100% vorkommen), während andere sogar sich mitziehen lassen von der Perturbation. Letzteres trifft auf drei Sprecherinnen zu wie Tabelle 3.3 die die Kompensationsstärke pro Sprecher auflistet, zeigt.

Abbildung 3.21 zeigt die nach Formel 3.2 berechneten Adaptive Response-Werte für *f0* über alle Sprecher hinweg. Ein Lineares Gemischtes Modell mit den Adaptive Response-Werten für *f0* als abhängiger Variable, und den Experimentphasen (mit den Stufen *BASELINE, PERTURBATION1, PERTURBATION2, RÜCK, PERTURBATION3, PERTURBATION4*) sowie den Sprechern als Fehler ergab einen signifikanten Effekt von *Experimentphase* ($\chi^2[5] = 157,9; p < 0,001$); post-hoc durchgeführte Tukey-Tests ergaben signifikante

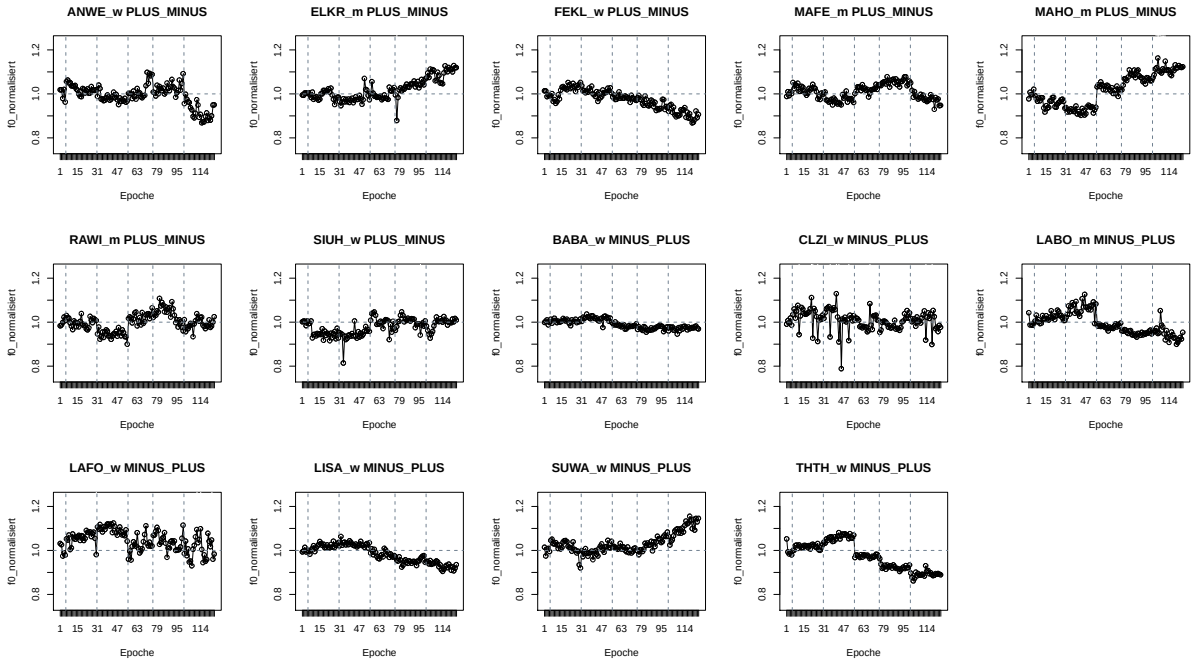

Abbildung 3.19: *Die durchschnittlichen, produzierten Grundfrequenzwerte der 14 Sprecher in zur Baseline normalisierter Form. Die ersten sieben Versuchspersonen gehören zur PLUS-MINUS-Gruppe, die nächsten sieben zur MINUS-PLUS-Gruppe. Die vertikalen Linien zeigen die Grenzen zwischen den Perturbationsphasen: der BASELINE-Phase folgen die Perturbationsphasen 1 (f0 um einen Halbton verschoben) und 2 (f0 um zwei Halbtöne verschoben), wobei bei den Versuchspersonen der MINUS-PLUS-Gruppe in Richtung tieferer Werte, bei jenen der PLUS-MINUS-Gruppe nach höheren Werten hin perturbiert wurde. Der daran anschließenden vierten Phase, der RÜCK-Phase, folgen die Perturbationsphasen 3 (ein Halbton) und 4 (2 Halbtöne), wobei die f0-Verschiebung je nach Gruppe nun in die Gegenrichtung zu den Verschiebungen in den ersten beiden Perturbationsphasen erfolgt. Das Geschlecht ist in den Teilabbildungsüberschriften kodiert, ebenso wie die Versuchspersonenkürzel.*

Abweichungen der Werte innerhalb der *BASELINE* und denen der *PERTURBATION1*- ($z = 2,9; p < 0,05$), *PERTURBATION2*- ($z = 7,2; p < 0,001$) und *PERTURBATION3*-Phase ($z = 4,3; p < 0,001$), nicht aber zwischen *BASELINE* und *PERTURBATION4* ($z = 2,2; n.s.$). Die Vergleiche mit der *RÜCK*-Phase zeigen ein sehr ähnliches Bild (zu PERTURBATION1: $4,2; p < 0,001$, PERTURBATION2: $z = 11,4; p < 0,001$, PERTURBATION3: $z = 6,6; p < 0,001$, PERTURBATION4: $z = 3,1; p < 0,05$); die Vergleiche zwischen *PERTURBATION1* und *PERTURBATION2* ($z = 7,6; p < 0,001$) bzw. zwischen *PERTURBATION3* und *PERTURBATION4* ($z = 3,7; p < 0,005$) erwiesen sich als signifikant. Das Bild, das Abbildung 3.21 zeigt, wird also bestätigt: in den ersten beiden Perturbationsstufen steigt Adaptive Response von *f0* treppenstufig signifikant an (d.h. also auch, dass die Perturbationsstärke eine Rolle gespielt hat), fällt wieder zur *Rück*-Phase

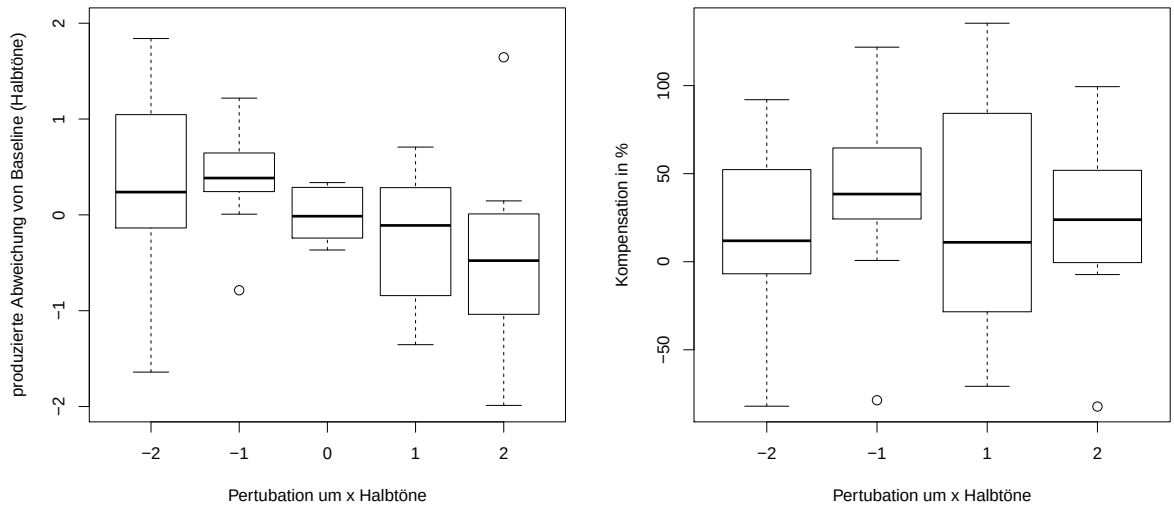

Abbildung 3.20: *Kompensation für Grundfrequenzperturbation, links dargestellt in Halbtönen, aufgeteilt in die Perturbationsstärken (−2 bis +2 Halbtöne). Rechts: Kompensation für Perturbation in prozentualen Anteilen. Eine ANOVA mit Messwiederholung ergab für diese Daten keinen Haupteffekt ($F[3, 39] = 0,85; n.s.$)*

SprecherIn	**Kompensation in %**	**SprecherIn**	**Kompensation in %**
SUWA	-27.7	FEKL	-48.4
CLZI	17.9	ANWE	-13.5
BABA	24.5	MAFE	17.4
LAFO	24.7	RAWI	27.6
LISA	50.3	SIUH	34.1
LABO	55.7	ELKR	38.7
THTH	78.2	MAHO	87.2

Tabelle 3.3: *Sprecher, Kompensation in Prozent, und Sprechergruppe Links: die Versuchspersonen der* MINUS-PLUS-*Gruppe. Rechts: die Versuchspersonen der* PLUS-MINUS-*Gruppe. Die Werte sind nach Kompensationserfolg geordnet, oben beginnend mit den niedrigsten Werten.*

hin ab (die sich von der *BASELINE*-Phase in diesem Experiment kaum unterscheidet: $z = 0,27; n.s.$), um zur *PERTURBATION3*-Phase wieder deutlich anzusteigen (auf etwas die Werte in *PERTURBATION1*, was der Vergleich dieser Phasen zeigte ($z = 2,5; n.s.$)). In der letzten Phase wird im Durchschnitt allerdings wieder weniger kompensiert, was, wenn man Abbildung 3.20 betrachtet, allerdings offenbar auf den Einfluss einiger weniger Sprecher zurückzuführen sein dürfte.

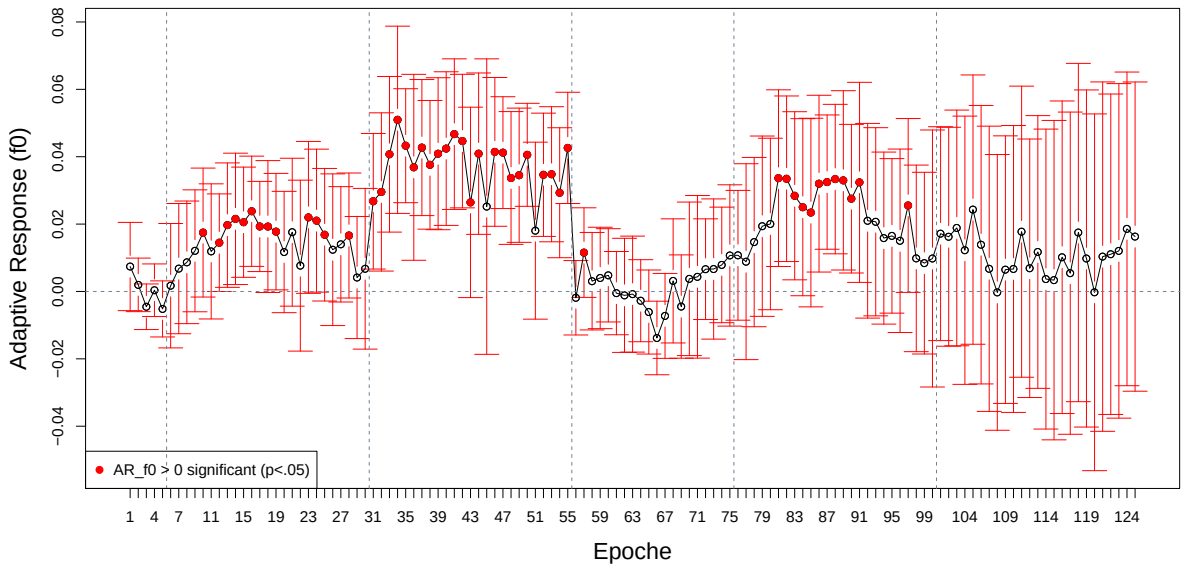

Abbildung 3.21: *Adaptive-Response-Werte für f0 der 14 Sprecher (siehe Formel 3.2). Die Abbildung repräsentiert einen Wert pro Sprecher pro Epoche, wobei die t-Verteilung der Sprecherwerte pro Epoche als rote Balken gezeigt werden. Der Kreis innerhalb dieser Verteilung entspricht dem arithmetischen Mittel und ist dann rot ausgefüllt, wenn einseitige Einstichproben-t-Tests ergeben haben, dass die Werte für die gegebene Epoche signifikant größer als 0 sind.*

Abbildung 3.22 zeigt die Adaptive Response -Werte für *F1*. Ein vergleichbares Lineares Gemischtes Modell, diesmal mit Adaptive Response von *F1* als abhängiger Variable durchgeführt, zeigte auch hier einen signifikanten Effekt für *Experimentphase* ($\chi^2[5] = 23,4; p < 0,001$). Die post-hoc Tukey-Tests offenbarten allerdings, dass es keine signifikanten Unterschiede zwischen der *BASELINE* und den vier Perturbationsphasen gab (zu PERTURBATION1: $z = 0,94; n.s.$, PERTURBATION2: $z = 0,69; n.s.$, PERTURBATION3: $z = 0,0; n.s.$, PERTURBATION4: $z = 0,77; n.s.$); der Vergleich der Perturbationsphasen mit der *RÜCK*-Phase allerdings brachte signifikante Ergebnisse für zwei Paarungen: *RÜCK* zu *PERTURBATION2* ($z = 3,76; p < 0,005$) und *RÜCK* zu *PERTURBATION4* ($z = 3,89; p < 0,005$); die anderen beiden Paarungen waren nicht signifikant (PERTURBATION1: $z = 1,08; n.s.$ und PERTURBATION3: $z = 2,63; n.s.$). Es gibt einen gerade

noch signifikanten Effekt im Vergleich zwischen *PERTURBATION2* und *PERTURBATION1* ($z = 2,84; p < 0,05$), aber keinen zwischen den Phasen 3 und 4 ($z = 1,34; n.s.$).

Wenn man also die *RÜCK*-Phase mit berücksichtigt, ergibt sich ein leichter Effekt, dass *F1* steigt. Dies ergibt sich auch, wenn man wiederum ein Lineares Gemischtes Modell rechnet, diesmal mit einer auf zwei Stufen (*BASIS*(=BASELINE+RÜCK) und *PERTURBATION*(Perturbationsphasen 1 bis 4))reduzierten unabhängigen Variable *PERTURBATION*. Ein solches Modell ergibt (für Adaptive Response von *f0* als abhängiger Variable) einen signifikanten Unterschied zwischen den beiden Phasen ($z = 9,57; p < 0,005$), und zu den Koeffizienten zählt eine leicht positive Steigung (0.007), wenn man eine Steigung für die Daten aller Sprecher berechnet (berechnet man eine Steigung pro Versuchsperson, ergibt sich ein Verhältnis von 5 negativen zu 9 positiven). Es scheint also der Fall zu sein, dass der erste Formant leicht steigt, wenn die Grundfrequenz steigt.

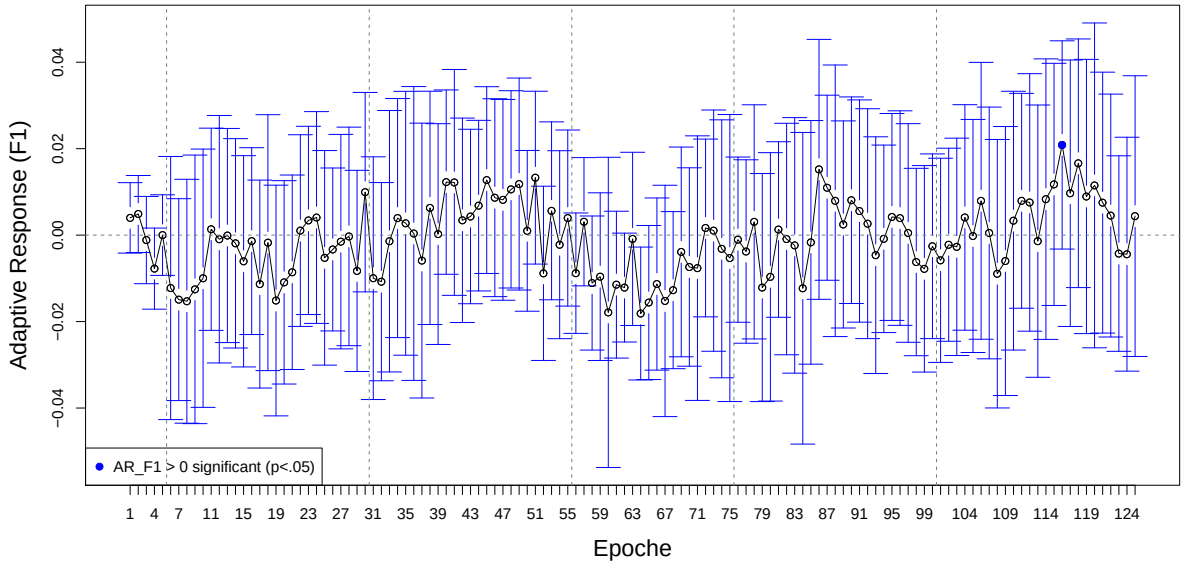

Abbildung 3.22: *Adaptive-Response-Werte für F1 der 14 Sprecher (siehe Formel 3.2). Die Abbildung repräsentiert einen Wert pro Sprecher pro Epoche, wobei die t-Verteilung der Sprecherwerte pro Epoche als blaue Balken gezeigt werden. Der Kreis innerhalb dieser Verteilung entspricht dem arithmetischen Mittel und ist dann blau ausgefüllt, wenn einseitige Einstichproben-t-Tests ergeben haben, dass die Werte für die gegebene Epoche signifikant größer als 0 sind.*

Diese Korrelation wurde mittels linearer Regressionsanalyse beschrieben. Hierzu wurde erneut die Funktion *lmer* benutzt , um die Abhängigkeit der Adaptive Response-Werte des ersten Formanten von den Adaptive Response-Werten der Grundfrequenz zu überprüfen, unter Ausklammerung der sprecherbedingten Variation. Das Modell ergibt einen signifikanten Effekt ($\chi^2[1] = 172,23; p < 0,001$) und eine positive Steigung (0.30). Wird pro

Sprecher eine Steigung errechnet, ergibt sich für nur drei Sprecher eine negative Steigung (SUWA, MAFE und ELKR), für die übrigen 11 Sprecher jedoch eine positive.

Um festzustellen, ob bezüglich dieser positiven Korrelation der Adaptive Response-Werte von *F1* und *f0* Unterschiede bestehen, die mit der Perturbation oder Nicht-Perturbation zusammenhängen, musste eine Kovarianzanalyse mit Messwiederholung, also eine Verbindung von linearer Regressionsanalyse und Varianzanalyse und gleichzeitiger Ausklammerung unerwünschter Variationsbringer, betrieben werden; auch dies kann man mit der Implementierung Linearer Gemischter Modelle im *R*-package *lme4* und seiner Funktion *lmer* bewerkstelligen. Die Sprecher werden ausgeklammert, und die abhängige Variable *Adaptive Response* von *F1* mit zwei unabhängiger Variablen modelliert, nämlich *Adaptive Response* von *f0* und *PERTURBATION* (mit zwei Stufen: *BASIS* und *PERTURBATION*). Dieses Modell ergab eine signifikante Interaktion zwischen beiden unabhängigen Variablen ($z = 15,4; p < 0,001$). Lineare Regressionen zwischen den Adaptive Response-Werten von *F1* und *f0*, getrennt durchgeführt für die Daten aus den *BASIS*-Phasen und jenen aus der Perturbationsphasen, ergab eine mittlere Steigung von 0.46 für die *BASIS*-Daten und von 0.26 für die *PERTURBATION*-Daten. Wenn man allerdings jeweils Steigungen pro Sprecher errechnet und diese einem gepaarten t-Test zuführt (gepaart deswegen, da jeweils pro Sprecher ein Wert für *BASIS* und ein Wert für *PERTURBATION* errechnet wurde), zeigt sich, dass es keinen konsistenten und signifikanten Unterschied gibt ($t[13] = 0,59; n.s.$). Abbildung 3.23 zeigt die Verteilung der ermittelten Steigungen über alle Sprecher - wie man sieht, steigt die Varianz unter Perturbation stark an, d.h. bei einigen Sprechern wird die Korrelation zwischen den Adaptive Response-Werten von *F1* und *f0* sogar stärker, bei einigen Sprecher sinkt sie und bei vier Sprechern wird sie gar negativ (SUWA, MAFE, LISA, und ELKR). Freilich gilt zu beachten, das eine Steigung um 0 bedeutet, dass kein Zusammenhang zwischen den Maßen besteht.

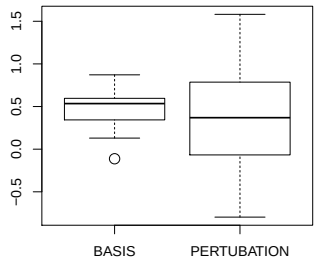

Abbildung 3.23: *f0-Perturbation: Verteilung der Steigungen in der pro Sprecher errechneten Regression von Adaptive Response(F1) Adaptive Response(f0).*

Halten wir also fest, dass eine gewisse Tendenz zu einer Korrelation der Adaptive Response-Werten von *F1* und *f0* besteht, also *F1* steigt, wenn *f0* steigt (siehe auch Abbildung 3.24 auf Seite 136). Dies ist auch der Fall für die normierten *F1*- und *f0*-Werte, wie

die Abbildungen 3.27 auf Seite 137 zeigt. Wenn dies der Fall ist, muss überprüft werden, ob dies auch für höhere Formanten, d.h. in diesem Fall für *F2* und *F3* der Fall ist.

Adaptive Response-Werte von *F2* wurden mittels *lmer* mit Adaptive Response von *f0* modelliert, und zwar natürlich unter Ausklammerung der Variabilität zwischen den Sprechern. Ein leicht signifikanter Effekt wurde hierbei festgestellt ($\chi^2[1] = 6,12; p < 0,05$), mit leicht negativer Steigung (-0.07). Da bezüglich Vokalhöhe und *F2* vernünftigerweise keine Voraussagen getroffen werden können, ist es sinnvoller, nicht das abstraktere Maß der Adaptive Response zu verwenden, sondern die zur *BASELINE* normalisierten Werte. Eine vergleichbare Modellierung, mit den normierten *F2*-Werten als abhängiger und den normierten *f0*-Werten als unabhängiger Variable ergab keine Signifikanz ($\chi^2[1] = 1,08; n.s.$).

Für *F3* ist das Bild das folgende: Weder bei Verwendung der Adaptive Response-Werte ($\chi^2[1] = 3,84; n.s.$) noch bei Verwendung der normalisierten Werte ($\chi^2[1] = 0,20; n.s.$) ergeben sich Signifikanzen.

Zur Beantwortung, ob beide Parameter (*F2* und *F3*) durch den Faktor *PERTURBATION* (Stufen: *BASIS* vs. *PERTURBATION*) beeinflusst wurden, mussten wieder die Adaptive Response-Werte als abhängige Variable in einem Linearen Gemischten Modell herangezogen werden. Keiner der Parameter erwies sich aber in dieser Modellierung durch *PERTURBATION* beeinflusst (*F2*: $\chi^2[1] = 1,68; n.s.$; *F3*: $\chi^2[1] = 2,34; n.s.$).

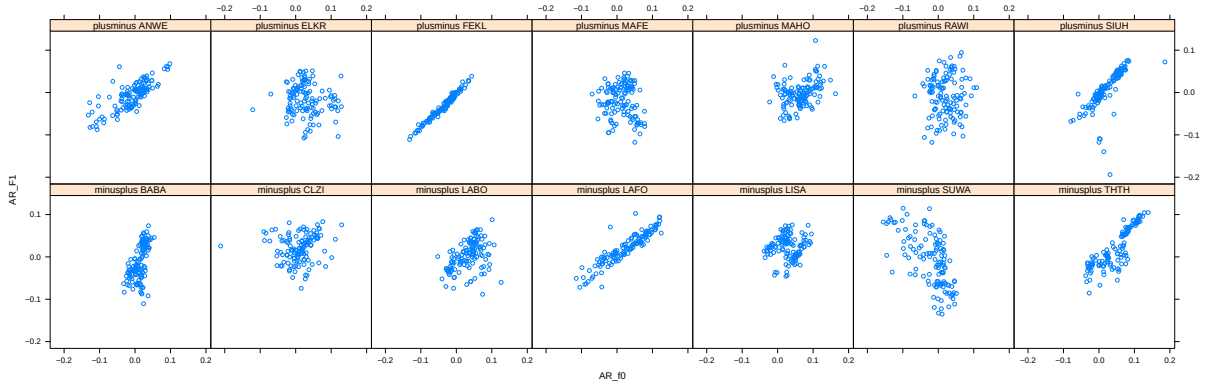

Abbildung 3.24: $AdaptiveResponse(F1) \sim AdaptiveResponse(f0)$.

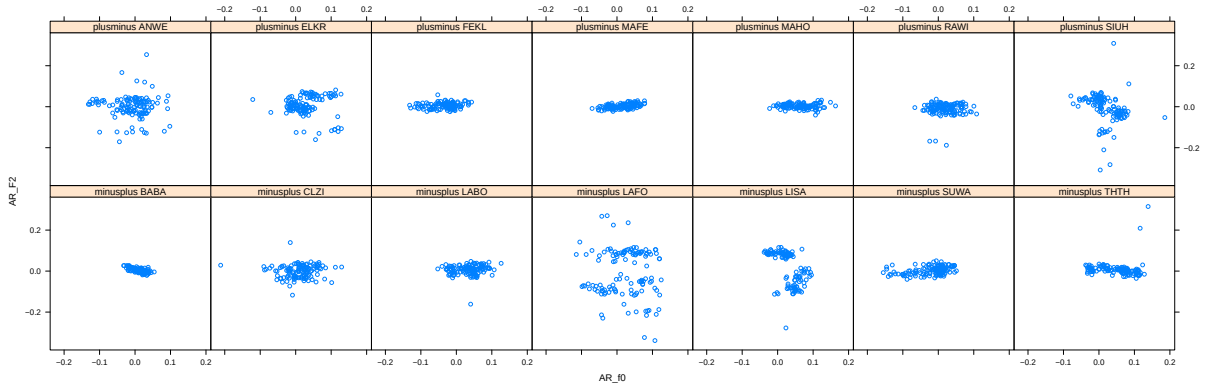

Abbildung 3.25: $AdaptiveResponse(F2) \sim AdaptiveResponse(f0)$.

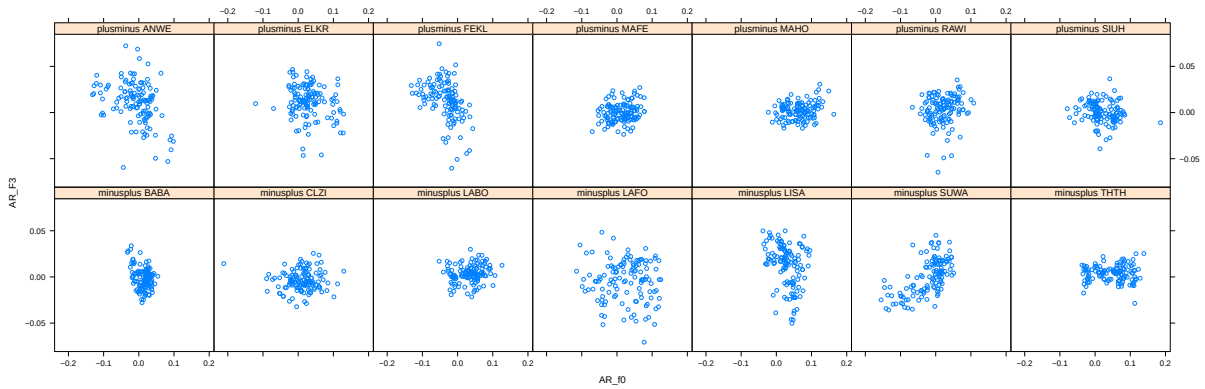

Abbildung 3.26: $AdaptiveResponse(F3) \sim AdaptiveResponse(f0)$.

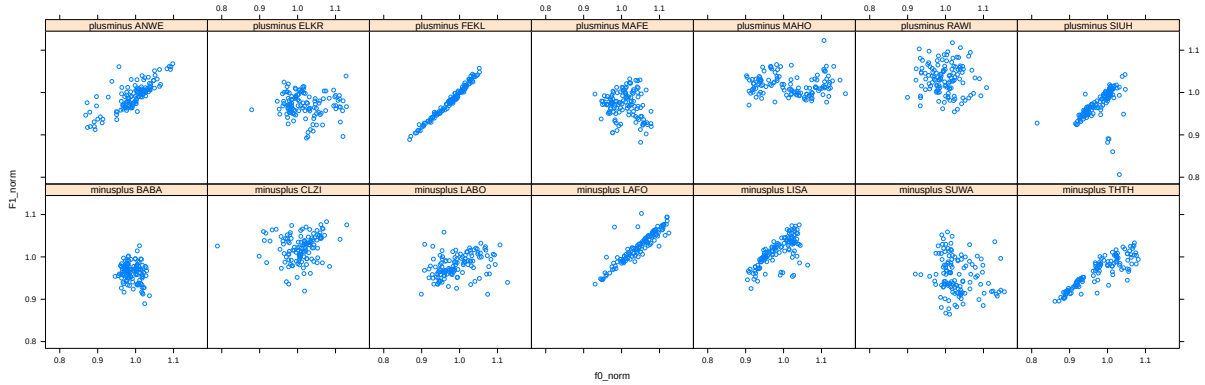

Abbildung 3.27: $F1(normalisiert) \sim f0(normalisiert)$.

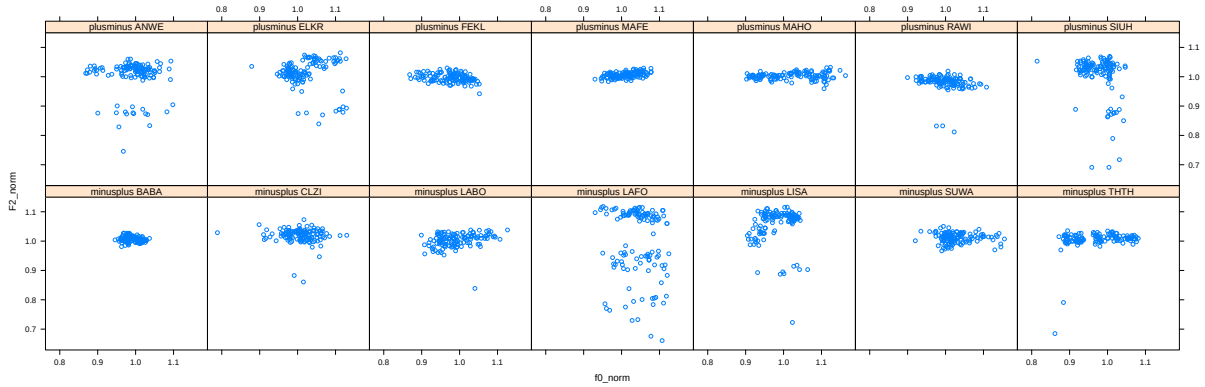

Abbildung 3.28: $F2(normalisiert) \sim f0(normalisiert)$.

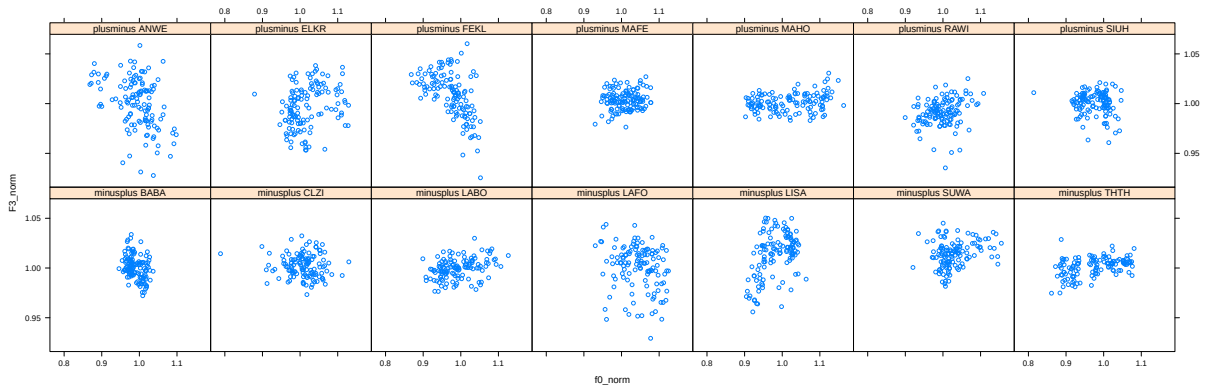

Abbildung 3.29: $F3(normalisiert) \sim f0(normalisiert)$.

Es darf hier nicht verschwiegen werden, dass einige Sprecherinnen (insbesondere ANWE, FEKL, LAFO, LISA, SIUH, und THTH) eine geradezu perfekt lineare *F1* ∼ *f0*-Korrelation aufweisen. Betrachten wir in der nächsten Abbildung deren Rohdaten (3.30), so müssen wir feststellen, dass ihre Werte nahezu perfekt $F1 = 2 \times f0$ abbilden; d.h. leider, dass möglicherweise die Ergebnisse für *F1* für diese Sprecherinnen durch die erste Harmonische beeinflusste Artefakte des Formantbestimmungsalgorithmus sein könnten, wie sie auch Iseli, Shue und Alwan (2006) beschreiben.

Wiederholen wir also die Statistik bezüglich *F1* ohne diese sechs Sprecherinnen, so ergibt sich in einem Linearen Gemischten Modell mit der Adaptive Response von *F1* als abhängiger Variable, der unabhängigen Variable *PERTURBATION* (mit den zwei Stufen *BASIS* und *PERTURBATION*) sowie unter Ausklammerung der Variation, die durch die nunmehr nur noch acht Sprecher eingebracht wird, kein Effekt mehr ($\chi^2[1] = 0,99; n.s.$), sehr wohl allerdings in den Daten der anderen sechs Sprecherinnen ($\chi^2[1] = 14,23; p < 0,001$). Auch, wenn man Adaptive Response von *F2* ($\chi^2[1] = 0,0027; n.s.$) und *F3* ($\chi^2[1] = 3,13; n.s.$) als abhängige Variablen benutzt, ergeben sich keine signifikanten Beeinflussungen durch *PERTURBATION* für diese acht Sprecher, doch dies ist auch nicht für die sechs Sprecherinnen mit den vermuteten Artefakten der Fall (*F2*: $\chi^2[1] = 2,06; n.s.$, *F3*: $\chi^2[1] = 0,27; n.s.$) - d.h. sollte es sich bei dem Ergebnis für *F1* bei diesen sechs Sprecherinnen um ein Artefakt handeln, betrifft diese Beeinflussung durch die Grundfrequenz offenbar nur die Messung des ersten Formanten.

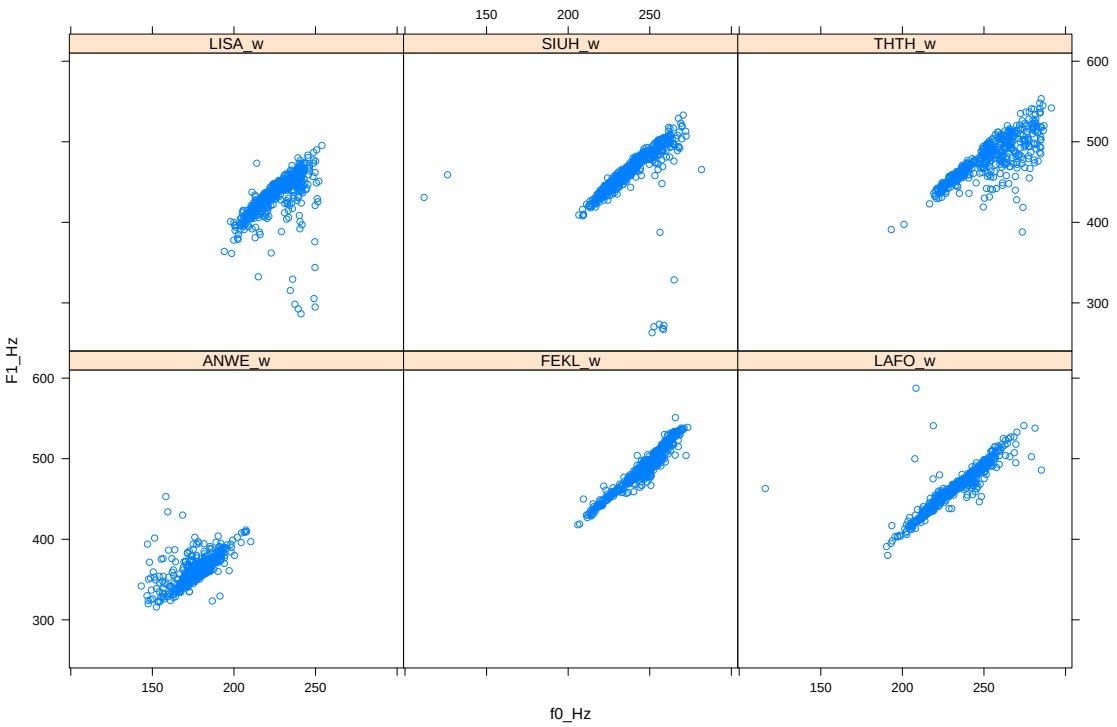

Abbildung 3.30: $F1(Hz) \sim f0(Hz)$.

Post-hoc Test 1: Hätte ein anderer Formantextraktionsalgorithmus andere Ergebnisse geliefert?

Es geht über den Themenbereich dieser Dissertation hinaus, die korrekte Funktion verschiedener Formantextraktionsalgorithmen zu evaluieren, siehe hierzu z. B. Vallabha und Tuller (2002) oder Wempe und Boersma (2003). Wir wollen aber, da bei den genannten sechs Sprecherinnen sehr auffällige Zusammenhänge zwischen den *f0*- und den *F1*-Maßen bestehen (die vergleichbar sind mit jenen, ebenfalls bei Sprecherinnen mit hoher Grundfrequenz gefundenen in Iseli et al. (2006), wo die *F1*-Werte sehr nahe an ganzzahligen Vielfachen der Grundfrequenz lagen), überprüfen, ob wir dies hätten vermeiden können, indem wir einen anderen, standardmäßig benutzten Formantextraktionsalgorithmus verwendet hätten. Der bislang in dieser Arbeit benutzte, wie er in *forest* implementiert ist, wird nicht nur am Institut für Phonetik und Sprachverarbeitung der LMU München standardmäßig benutzt. Wie vergleichen diesen mit dem Standard-Formantextraktionsalgorithmus in dem weit verbreiteten Programm *praat*, dem *Burg*-Algorithmus (vergleiche (Childers, 1978, Seiten 252-255) und (Press, Teukolsky, Vetterling & Flannery, 1992, Seiten 568-570)).

Die auffälligste *F1* $\sim$ *f0*-Korrelation ist bei Sprecherin FEKL zu finden. Wir messen für diese Sprecherin mit dem erwähnten *Burg*-Algorithmus und wählen hierzu die Standardeinstellungen für weibliche Sprecherinnen, also eine Fensterverschiebung von 5 ms, 5 als die maximale Anzahl der zu findenden Formanten, 5500 Hz als Obergrenze des Suchraums, und eine Messfensterlänge von 25 ms sowie Präemphase ab 50 Hz. Diesen Messungen wird für jeden Vollvokal ein Wert zum zeitlichen Mittelpunkt entnommen. So entsteht ein Vektor mit Messwerten, der parallel zu demjenigen ist, der bislang zur Analyse benutzt wurde; dies bedeutet, dass für jeden der 500 /eː/-Vokale aus dem Grundfrequenzperturbationsexperiment für Sprecherin FEKL nun zwei Werte für F1 existieren: einer wurde durch *forest* gewonnen, einer durch *praat*s *Burg*-Methode.

Zu beachten ist hierbei, dass sich nicht allein die Algorithmen unterscheiden, sondern auch die Vorgehensweise bei der Messung. In der bislang benutzten Methode wurde der Medianwert aus den mittleren 20 % der Vokale benutzt, hier der Wert zum zeitlichen Mittelpunkt. Der bislang benutzte Medianwert entsprang Messungen, die jede Millisekunde stattfanden, um es der üblichen Schrittweite im Grundfrequenzextraktionsalgorithmus von *STRAIGHT* gleichzutun.

Vergleicht man die Werte aus *forest* und *praat* mittels eines gepaarten t-Tests, ergibt sich zwar eine signifikante Abweichung ($t[499] = 13,7; p < 0,001$), die mittlere Differenz beträgt aber nur 2.378 Hz, d.h. *praat* liefert im Durchschnitt minimal höhere Werte.

Korreliert man diese Werte miteinander, wie in Abbildung 3.31, so ergibt sich eine große Übereinstimmung der Werte; der adjustierte R^2-Wert beträgt 0.9782 (p<0,001), also fast 1, und die Steigung ist dementsprechend auch nahe 1 (1.028).

Der Vergleich beider Standard-Formantextraktionsalgorithmen zeigt also, dass auch die Verwendung des *Burg*-Algorithmus, wie er in *praat* zu finden ist, wohl nicht zu anderen Ergebnissen für F1 geführt hätte. Sollte der enge Zusammenhang zwischen der Grundfrequenz und dem ersten Formanten wirklich ein Artefakt der Messung sein, beeinflusst durch die erste Harmonische der Grundfrequenz, so wären auch im *Burg*-Algorithmus die

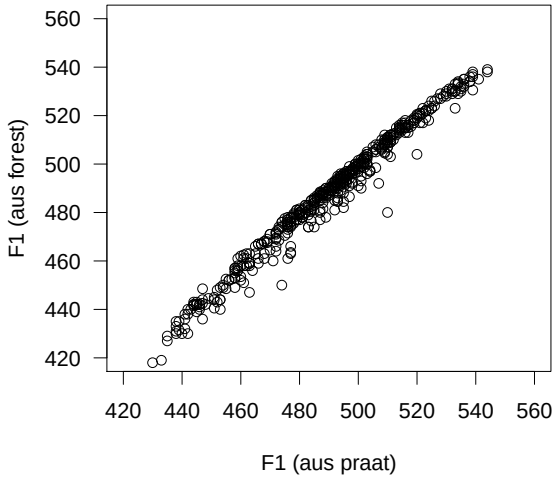

Abbildung 3.31: *F1 (aus* forest*)* ∼ *F1 (aus* praat*) bei Sprecherin FEKL.*

F1-Werte durch *f0* beeinflusst worden.

Der Zweitgutachter der vorliegenden Arbeit hatte in seinem Gutachten darauf hingewiesen, man könne „durchaus daran denken, die Formantanalyse im Formantmanipulationsprogramm (transshift von Shanqing Cai) anzuwenden, weil hier mit einer Glättung im Cepstralbereich versucht wird, die typischen LPC-Probleme bei weiblichen Probanden zu minimieren“. Gleichzeitig müsse man versuchen, „die LPC-Parameter für diese individuelle Stimme zu optimieren, was bei solchen Problemfällen unbedingt erforderlich wäre.“

Aufgrund dieser Anregung wurde versucht, die für Sprecherin FEKL optimalen Parameter zu finden und gleichzeitig die erwähnte Glättung im Spektralbereich wie vorgeschlagen durchzuführen. Es wurden für die genannte Sprecherin eine Reihe von Parametereinstellungen ausprobiert und im Nachhinein durch Inspektion anhand der dem Spektrogramm überlagerten Formantwerte die offenbar am wenigsten fehleranfällige Einstellung ausgewählt (eine LPC-Ordnung von 19 bei einem frameshift von 20 ms). Leider erwies sich auch dieses sehr aufwändige Verfahren als nicht zielführend: es blieb weiterhin eine starke, positive Korrelation mit der Grundfrequenz erhalten, wenn auch die F1-Werte etwas niedriger errechnet wurden und somit kein einfacher Zusammenhang wie oben erwähnt (also mit $F1 = 2 * f0$) aufzuzeigen war. Ähnlich wie in dem Vergleich mit den in *praat* errechneten Werten ergibt sich bei Korrelation der neuen, in *transshift* ermittelten Werte mit jenen aus *forest* eine Steigung nahe 1 (0.96). Da es aussichtslos erscheint, weitere, ”korrektere”Messungen vorzunehmen, enden hiermit die Versuche in dieser Richtung.

Post-hoc-Test 2: Spielen Intensitätsunterschiede eine Rolle?

Es erscheint zwar wahrscheinlich, dass der vergleichsweise geringe Effekt der Perturbationsphasen auf den ersten Formanten im Grundfrequenzperturbationsexperiment auf Artefakte bei einigen Sprecherinnen zurückzuführen sein dürfte, doch ist eine tatsächliche Korrelation des ersten Formanten mit der Grundfrequenz doch nicht so unwahrscheinlich (siehe z. B. Syrdal und Steele (1985), Maurer, Landis und D'heureuse (1991); Maurer, Cook et al. (1991); Maurer und Landis (1995) oder auch Chládková et al. (2009)), dass man nicht mögliche Kovariablen testen sollte.

Liénard und Di Benedetto (1999) und Eriksson und Traunmüller (2002) zeigten, dass sowohl *f0* als auch *F1* steigen, wenn der sogenannte *vocal effort* erhöht wird, also das, was sich als perzebierbare Korrelate zur Distanz zwischen Sprecher und Hörer äußert. Dementsprechend ist es vorstellbar, dass die Sprecher in den hier vorgestellten Experimenten möglicherweise durch die Perturbation veranlasst mit erhöhter Intensität sprachen, und es möglicherweise deshalb zu einem gleichzeitigen Anstieg von Grundfrequenz und erstem Formanten kam. Um die Intensität zu messen, verwenden wir hier die Kurzzeit Root-mean square (rms)-Amplitude als Korrelat der lokalen Energie eines Signals und damit der Intensität der Sprachsignale.

Um rms zu messen, wurde in *emu* die Funktion rmsana mit den Default-Einstellungen benutzt. In *emu-R* wurden diese Daten eingelesen und weiterbearbeitet. Für jedes /eː/-Token wurden die Werte extrahiert und der Medianwert errechnet. So ergibt sich für jedes /eː/ ein Wert in Dezibel[10].

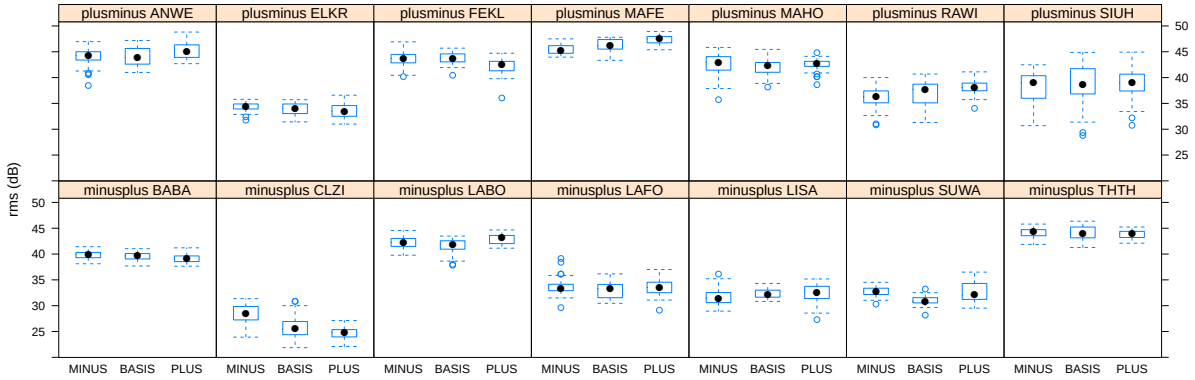

Abbildung 3.32: *Root-mean square (rms)-Werte pro Sprecher und pro Stufe des Faktors PERTURBATIONSRICHTUNG (BASIS=Keine Perturbation, MINUS, PLUS) im f0-Perturbationsexperiment*

Ein Lineares Gemischtes Modell, gerechnet mit der abhängigen Variable *rms*, und mit

[10]Wegen der besseren Vergleichbarkeit zwischen den Sprechern und wegen der teilweise nötigen gelegentlichen Anpassung des Aufnahmepegels wurden alls rms-Werte normalisiert, indem den vier Werten pro Aufnahme das arithmetische Mittel der rms-Werte aus den zwei Pausen-Segmenten (eine vor und eine nach der Äußerung) der gleichen Aufnahme subtrahiert wurde.

dem dreistufigen Faktor *PERTURBATIONSRICHTUNG* (Stufen: (*MINUS*, *BASIS* (=keine Perturbation), *PLUS*)) als unabhängiger Variable, ergab, unter Ausklammerung der sprecherbedingten Variation, einen signifikanten Effekt ($\chi^2[2] = 9,37; p < 0,01$) der *PERTURBATIONSRICHTUNG*. Post-hoc durchgeführte Tukey-Tests zeigten aber, dass die rms-Werte sowohl bei *MINUS*($z = 2,81; p < 0,05$) als auch bei *PLUS*($z = 2,78; p < 0,05$) signifikant höher waren als bei *BASIS*, während sie sich untereinander nicht unterschieden ($z = 0,033; n.s.$). Es ist also die Perturbation an sich, die die rms-Werte nach oben treibt, unabhängig jedoch von der *PERTURBATIONSRICHTUNG*.

Die oben genannte Modellierung wurde anschließend getrennt für die Sprecherinnen, die die starke *F1* $\sim$ *f0*-Korrelation aufweisen (also für ANWE, FEKL, LAFO, LISA, SIUH, und THTH), und für die anderen Sprecher (ELKR, MAFE, MAHO, RAWI, BABA, CLZI, LABO und SUWA) durchgeführt. Gerade für die sechs Sprecherinnen ergibt sich jedoch kein konsistenter und signifikanter Effekt des Faktors *PERTURBATIONSRICHTUNG*($\chi^2[2] = 0,31; n.s.$), wohl aber für die anderen acht Sprecher ($\chi^2[2] = 18,04; p < 0,001$), wobei bei diesen wiederum die Perturbation an sich eine Rolle spielt, und nicht die Richtung derselben (*BASIS* - *MINUS*: $z = -4,094; p < 0,001$, *PLUS* - *BASIS*: $z = 3,61; p < 0,001$, *PLUS* - *MINUS*: $z = -0,59; n.s.$).

Für die MINUS-Perturbation, bei der die Sprecher *f0* erhöhen, ist also rms höher (wie erwartet), aber eben auch bei der PLUS-Perturbation, bei der die Sprecher *f0* in der Mehrzahl absenken; außerdem ist gerade bei den Sprecherinnen, bei denen zumindest den Messwerten nach *F1* kovariiert, kein Effekt für rms zu finden. Die RMS-Werte an sich können also auch nicht erklären, warum *F1* und *f0* positiv miteinander korreliert sein sollten.

Aus Vergleichsgründen wurde auch für die Daten aus dem *F1*-Perturbationsexperiment rms gemessen und ausgewertet. Abbildung 3.33 zeigt schon, dass es hier einen eindeutigeren Zusammenhang zwischen Perturbationsrichtung und rms-Werten gibt, obschon dort ja in der Regel *F1* und *f0* in gegensätzliche Richtungen kompensatorisch verschoben wird. Die Statistik, ausgeführt wie oben, ergibt einen stark signifikanten Effekt der Perturbationsrichtung auf die rms-Werte ($\chi^2[2] = 434,1, p < 0,001$). Wie die Abbildung schon vermuten lässt, sind die Werte der rms bei *MINUS* gegenüber *BASIS* in der Regel höher (Tukey-Test: $z = 8,14; p < 0,001$), die bei *PLUS* niedriger ($z = 9,97; p < 0,001$). Dies kann recht einfach erklärt werden, da bei der MINUS-Perturbation ja in der Regel der erste Formant kompensatorisch erhöht produziert wird (und damit der Öffnungsgrad steigt), und bei der PLUS-Perturbation der gleiche Effekt in umgekehrter Richtung auftritt, also der Öffnungsgrad sinkt.

Reanalyse der Daten von 8 Sprechern

Fasst man die Ergebnisse, die bislang beschrieben wurden, zusammen, stellt man fest, dass es eine scheinbare positive Korrelation zwischen *f0* und *F1* zu geben scheint, die aber vermutlich auf offensichtliche Artefakte zurückzuführen sind, dass rms-Werte hauptsächlich durch Perturbation beeinflusst werden und nicht durch die Perturbationsrichtung, und dass, wenn man die offensichtlich artefakt-behafteten Daten der erwähnten sechs Sprecherinnen ausschließt, und man die Daten der verbleibenden acht Versuchspersonen in Analogie zu

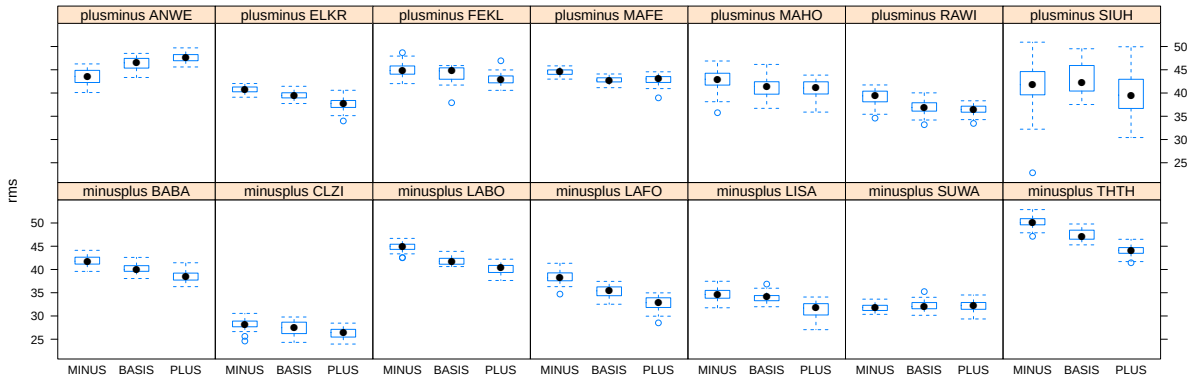

Abbildung 3.33: *Root-mean square (rms)-Werte pro Sprecher und pro Stufe des Faktors PERTURBATIONSRICHTUNG (BASIS=Keine Perturbation, MINUS, PLUS) im F1-Perturbationsexperiment*

Kapitel 3.2 mit den Adaptive Response-Werten als abhängiger Variable und mit dem Faktor *PERTURBATION* (*BASIS* vs. *PERTURBATION*) modelliert, man nur für *f0*, aber nicht für *F1, F2, F3* eine signifikante Beeinflussung findet.

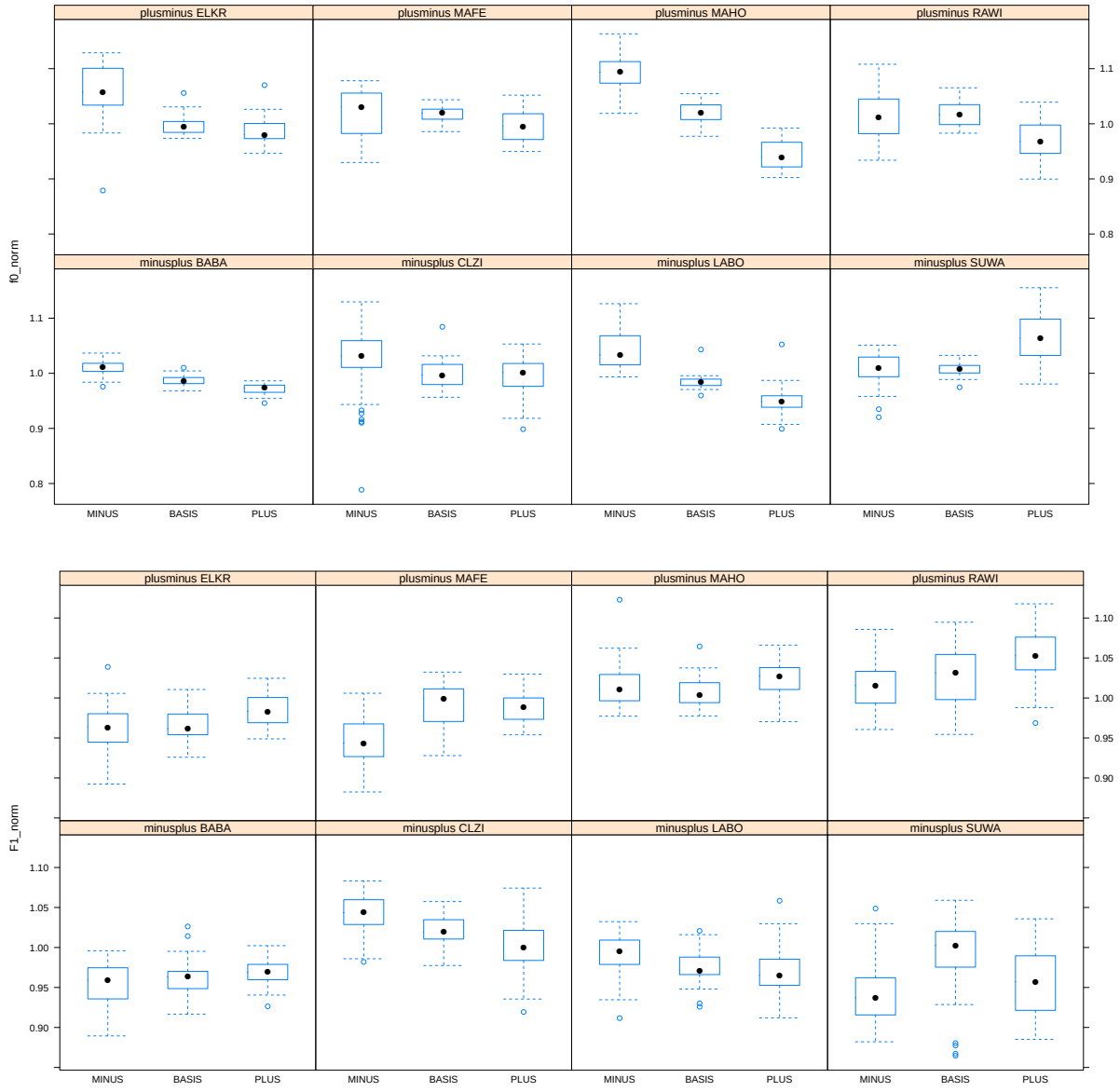

Abbildung 3.34: *Zur Baseline normierte f0- und F1-Werte pro Sprecher und pro Stufe des Faktors PERTURBATIONSRICHTUNG (BASIS=Keine Perturbation, MINUS, PLUS) im f0-Perturbationsexperiment*

Nun gilt es aber zu bedenken, dass die Grundfrequenz alleine sicher ein schwächerer auditorischer Cue zur Vokalhöhe ist als *F1*. Daraus folgend, wird in diesem Experiment möglicherweise das Vokalhöhenperzept der Sprecher weniger beeinflusst als im *F1*-Perturbationsexperiment. Sollten Vokalhöhenperturbationen durch die hier angewendeten *f0*-Perturbationen überhaupt zustande gekommen sein, so sind sie doch vermutlich nur schwach ausgeprägt, und könnten dementsprechend mit einer nur kleinen Korrektur des ers-

ten Formanten kompensiert werden; man muss also mit relativ geringen *F1*-Unterschieden rechnen. Daraus folgt, dass es gut möglich ist, dass die Unterschiede in den Adaptive Response-Werten zwischen den Stufen *BASIS* und *PERTURBATION* schlicht zu gering sind, um signifikant unterschiedlich zu sein. Dennoch ist es aber möglich, dass es Unterschiede zwischen den Perturbationsstufen *MINUS* und *PLUS* gibt, wenn man die zur *Baseline* normalisierten Werte verwendet. Hierbei geht zwar Information über die Reihenfolgeneffekte der Perturbationsstufen 1-4 verloren, was wir gerne verschmerzen wollen, zumal im Gegensatz zum *F1*-Perturbationsexperiment beim vorliegenden *f0*-Perturbationsexperiment keine allzu großen Reihenfolgeeffekte festzustellen sind, wie oben erwähnt wurde, da die Kompensation in der Regel mehr oder weniger sofort einsetzt.

Die Reanalyse der Daten der acht offenbar nicht von Messfehlern beeinträchtigten Versuchspersonen gestaltete sich folgendermaßen: in vier Linearen Gemischten Modellen wurde die abhängige Variable (die normierten Werte entweder von *f0*, *F1*, *F2* oder *F3*) modelliert durch die unabhängige Variable *PERTURBATIONSRICHTUNG* (Faktorstufen: *MINUS*, *BASIS* (=keine Perturbation), *PLUS*), und zwar, wie immer, unter Ausklammerung der Variation zwischen den Sprechern. Es ergaben sich die folgenden Ergebnisse:

- *f0*: Der Haupteffekt war signifikant $\chi^2[5] = 254,03; p < 0,001$. Post-hoc ergaben sich in einem Tukey-Test für alle drei möglichen Paar-Vergleiche signifikante Ergebnisse: *BASIS-MINUS*: $z = -8,5; p < 0,001$; *PLUS-MINUS*: $z = -16,9; p < 0,001$, *PLUS-BASIS*: $z = -5,3; p < 0,001$

- *F1*: Der Haupteffekt war signifikant $\chi^2[5] = 21,687; p < 0,001$. Post-hoc ergaben sich für zwei der drei möglichen Paar-Vergleiche signifikante Ergebnisse: *BASIS-MINUS*: $z = 3,14; p < 0,005$; *PLUS-MINUS*: $z = 4,45; p < 0,001$, *PLUS-BASIS*: $z = 0,49; n.s.$

- *F2*: Ein leichter Haupteffekt war zu finden $\chi^2[5] = 8,56; p < 0,05$. Post-hoc ergaben sich in einem der drei möglichen Paar-Vergleiche signifikante Ergebnisse: *BASIS-MINUS*: $z = -2,87; p < 0,05$; *PLUS-MINUS*: $z = -1,71; n.s.$, *PLUS-BASIS*: $z = 1,47; n.s.$

- *F3*: Ein signifikanter Haupteffekt war zu finden $\chi^2[5] = 25,99; p < 0,01$. Post-hoc ergaben sich in einem der drei möglichen Paar-Vergleiche signifikante Ergebnisse: *BASIS-MINUS*: $z = -2,91; p < 0,01$; *PLUS-MINUS*: $z = -5,04; p < 0,001$, *PLUS-BASIS*: $z = 1,21; n.s.$

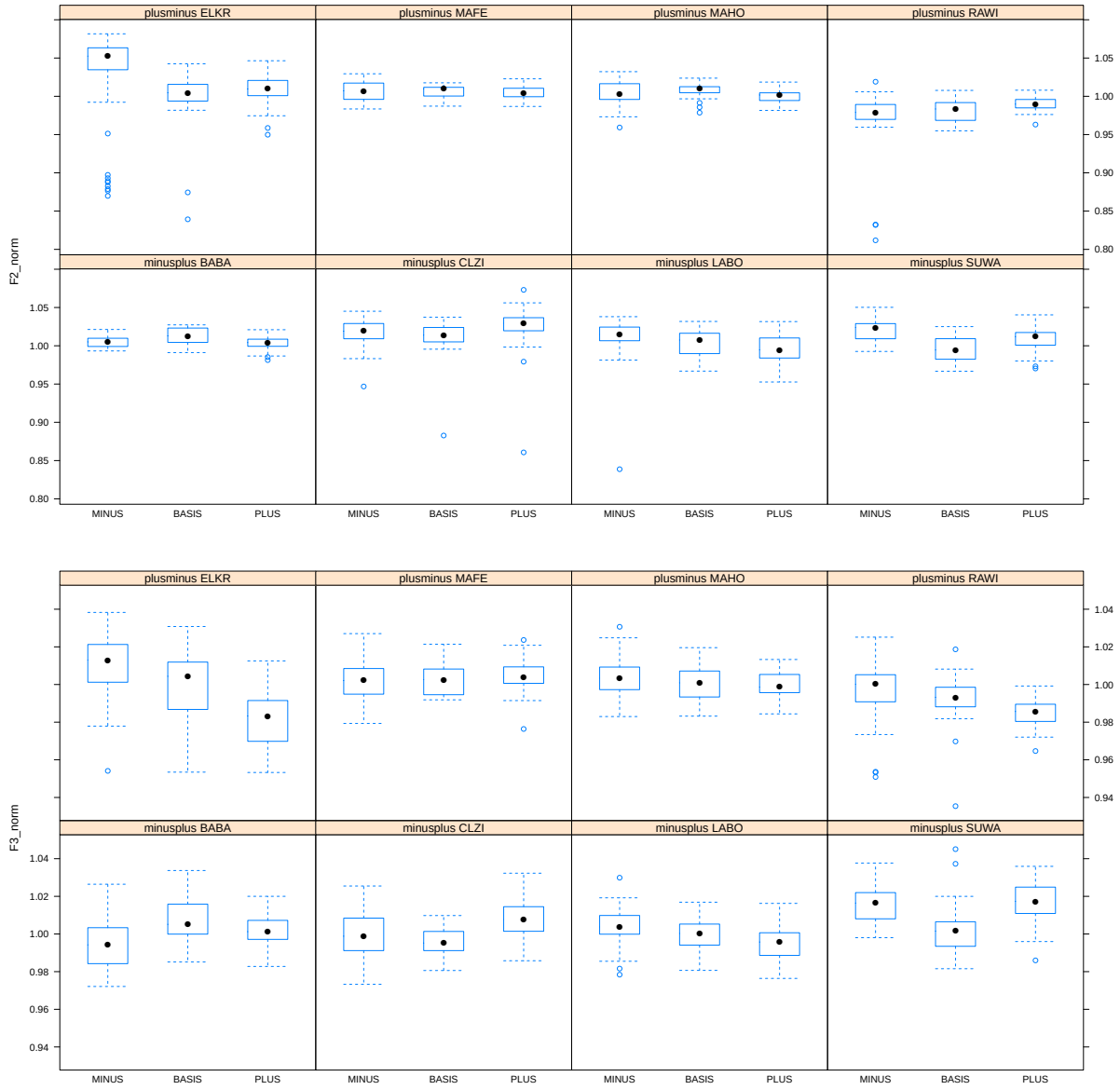

Abbildung 3.35: *Zur Baseline normierte F2- und F3-Werte pro Sprecher und pro Stufe des Faktors PERTURBATIONSRICHTUNG (BASIS=Keine Perturbation, MINUS, PLUS) im f0-Perturbationsexperiment*

In Analogie zu der Darstellung und der Statistik zu den Daten aus Abbildung 3.14 aus dem Formantperturbationsexperiment wollen wir pro Sprecher ein Wertepaar abbilden; im Gegensatz zur Vorgehensweise im *F1*-Perturbationsexperiment wollen wir nicht *BASIS* und *PERTURBATION* abbilden, sondern die *MINUS*-Phasen den *PLUS*-Phasen gegenüberstellen, ohne Unterscheidung der Perturbationsstärke. Wir wollen überprüfen, ob es wie im Formantperturbationsexperiment Steigungswerte gibt, die signifikant kleiner als

0 sind. Abbildung 3.36 zeigt die Daten für die acht Sprecher, die oben untersucht worden waren. Es gibt nur tendenziell negative Steigungen, und ein einseitiger t-Test (durchgeführt nach Überprüfung der Normalverteilung der Daten, $W = 0,94; n.s.$), der prüfte, ob die Werte kleiner 0 sind, ergab keine Signifikanz ($t[7] = -0,49; n.s.$). Dies relativiert die Ergebnisse des zuletzt ausgeführten Experiments.

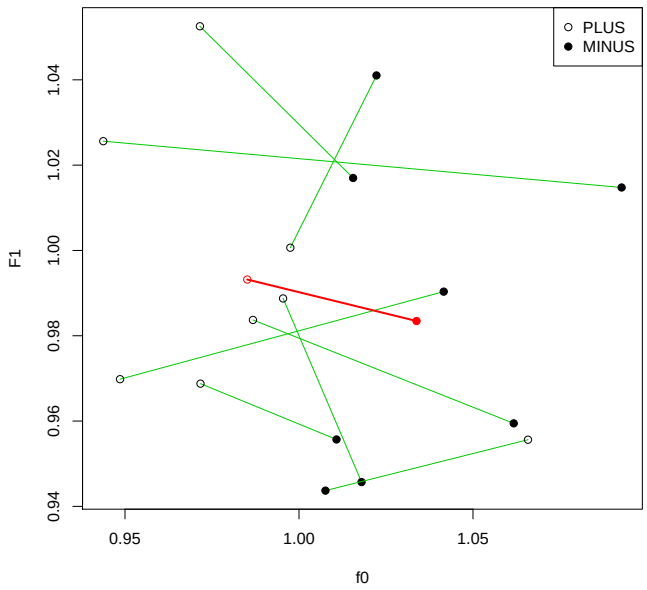

Abbildung 3.36: *f0-Perturbation. Die Mittelwerte über MINUS und PLUS Epochen der zur Baseline normierten Werte von* f0- *und* F1*; je ein Punktepaar pro Versuchsperson.*

3.3.3 Kurzzusammenfassung der Ergebnisse

Wir wollen hier noch einmal in aller Kürze die wichtigsten Ergebnisse zusammenfassen:

- Die erste Hypothese, dass Sprecher auf eine Perturbation der Grundfrequenz mit Veränderungen der Grundfrequenz kompensieren, aber zumindest teilweise unvollständig, konnte bestätigt werden
  - Sehr schnell nach Einführung der Grundfrequenzperturbation reagieren die Sprecher mit einer kompensatorischen Gegenbewegung, d.h. auch schon nach der sehr kurzen, 5 Epochen andauernden *Baseline*-Phase wird praktisch sofort nach Einsetzen der Perturbation kompensiert (ganz im Gegensatz zu den geringen Veränderungen im Formantperturbationsexperiment)
  - Je stärker die Perturbation, desto stärker die Veränderung der produzierten Grundfrequenz. Dies gilt zumindest für den Vergleich der Perturbationsphasen

1 und 2; außerdem unterscheiden sich die Ergebnisse der Perturbationsphasen 1 und 3, die jeweils die gleiche Perturbationsstärke von einem Halbton aufwiesen, nicht signifikant. Es gibt nicht den im Formantperturbationsexperiment beobachteten Effekt, dass die *Rück*-Phase sich von der *Baseline* unterscheiden würde. Abweichend von diesen erstaunlich schnellen und stufenweise voranschreitendem Kompensations- und Deadaptationsprozessen scheitert im Mittel die Kompensation in der letzten Perturbationsphase, wobei dies daher kommt, dass einige Sprecher hier zu *followers* werden, also der Perturbation folgen, während andere Sprecher erfolgreich kompensieren.

  - Es gibt in gewisser Weise drei Klassen von Sprechern in diesem Experiment: Ein Teil der Versuchspersonen kann sehr gut durch Kompensation in *f0* mit der *f0*-Perturbation umgehen; hierbei kompensieren einige wenige sogar nahezu vollständig (erste Klasse, zwei bis drei Sprecher, je nachdem, wo man die Grenze für nahezu vollständige Kompensation ziehen möchte), während andere - und zwar die Mehrheit der Sprecher - zumindest teilweise für die Störung kompensieren (zweite Klasse). Einige wenige verändern ihre Grundfrequenz in Richtung der Perturbation, kompensieren also überhaupt nicht (dritte Klasse, die *followers*). Unter der letzten Gruppe, die aus drei Sprecherinnen besteht, ist eine Sprecherin, die sich recht häufig versprach; möglicherweise hat sie die leichte Verzögerung, die durch die Addierung der Latenzzeiten von Grundfrequenzverschiebung und Mischpult zustandekommt (ca. 40 ms), wahrgenommen, und daher die Aufgabe nicht erfüllen können. Andererseits war die gleiche Versuchsperson auch im *F1*-Perturbationsexperiment ein *follower*.

- Die zweite Hypothese, dass der erste Formant sich *in Richtung* der Perturbation verschiebt, ist nicht eindeutig zu bestätigen; dennoch ist eine Tendenz dazu nicht zu leugnen

  - In den Perturbationsphasen ist die Zwischen-Sprechervariabilität sehr hoch. Über alle Sprecher gerechnet, ergeben sich signifikante Effekte nur, wenn man wie im Formantperturbationsexperiment die *Rück*-Phase mit berechnet. Dieser Effekt weist allerdings auf eine steigende Adaptive Response, also einen Anstieg der *F1*-Werte entgegen der Perturbation. Wie sich zeigt, kommt dies durch eine positive Korrelation der Grundfrequenz und des ersten Formanten zustande. Diese wiederum kommt wegen sechs Sprecherinnen zustande, deren gemessener erster Formant zumeist dem doppelten der Grundfrequenz entspricht. Wir vermuten hinter diesem Effekt ein Artefakt der Formantextraktion mittels LPC. Dieses Artefakt scheint aber nur den ersten Formanten zu betreffen, nicht *F2* oder *F3*, die ebenfalls ausgewertet wurden, um zu ermessen, ob eine Ansatzröhrlängenänderung durch Anhebung bzw. Absenkung des Kehlkopfes als eine Kovariable der Grundfrequenzänderung denkbar gewesen wäre.
  - Betrachtet man nur die anderen acht Versuchspersonen und vergleicht unperturbierte mit perturbierten Epochen, ergeben sich keine sigifikanten Ergebnisse,

sowohl nicht für *F1*, aber auch nicht für *F2* oder *F3*.

- Eine Reanalyse der Daten der acht Sprecher, die die *MINUS*-Perturbationsphasen, die *BASIS*-Phasen und die *PLUS*-Perturbationsphasen einander gegenüberstellt, zeigt zumindest Unterschiede zwischen den *MINUS*- und *PLUS*-Phasen: In den *PLUS*-Phasen, in denen *f0* kompensatorisch niedriger produziert wird als in der *MINUS*-Phase, ist der erste Formant größer als in der *MINUS*-Phase (aber nicht größer als in der *BASIS*-Phase); dort, wo *f0* erhöht wird (in der *MINUS*-Phase), ist der erste Formant erniedrigt (gegenüber der *PLUS*- als auch der *BASIS*-Phase). Zumindest der *MINUS-PLUS*-Vergleich zeigt also das erwartete Muster: wo *f0* steigt, sinkt *F1*, und umgekehrt.
- Eine weitere Analyse der Daten der acht Sprecher, über die ermittelt werden sollte, wie konsistent Sprecher das beschriebene Muster einer gegenläufigen Bewegung von Grundfrequenz und ersten Formanten anwendet, führte zu dem Ergebnis, dass der Gebrauch sehr inkonsistent ist. Da jedoch nur zwei Punkte für jeden der acht Sprecher für diese Statistik benutzt wurden, ist sicher die Aussagekraft dieser Analyse sehr eingeschränkt.
- Wegen dieser Beschränkung auf nur acht Personen verzichtet wir auch auf eine Analyse, ob *F1* dann bevorzugter verändert wird, je unvollständiger in *f0* kompensiert wird. Angemerkt sei allerdings, dass in der letztgenannten Analyse derjenige Sprecher, der am vollständigsten in *f0* kompensierte, und der also keinen Grund haben sollte, über das auditorische Feedback eine Vokalhöhenänderung wahrzunehmen, den geringsten Unterschied in *F1* zwischen der *MINUS*- und der *PLUS*-Phase produzierte, der nahezu gegen 0 tendierte.

3.4 Diskussion

In zwei Experimenten wollten wir feststellen, ob

- bei einer Vokalhöhenperturbation über eine quasi in Echtzeit durchgeführte Verschiebung des „klassischen“ Vokalhöhenkorrelats *F1* neben einer unvollständigen Kompensation in *F1* auch die Grundfrequenz von den Sprechern benutzt wird
- bei einer unvollständigen Kompensation für eine Grundfrequenzperturbation ebenfalls das Vokalhöhenperzept des Sprechers beeinflusst wird, was zu einer zusätzlichen Kompensation in *F1* führen sollte

Insbesondere die zweite Fragestellung ist entscheidend für diese Arbeit, da wir zeigen wollen, dass es vorstellbar ist, dass für langfristige, aber in ihrem Ausmaß beachtliche Veränderung der Grundfrequenz im Laufe des Lebens (bei den bei uns untersuchten Personen immerhin über 5 Halbtöne bei einem Mann und über 7 Halbtöne bei einer Frau) im Bereich des ersten Formanten kompensiert werden muss, um das Vokalhöhenperzept zu erhalten, dass durch die deutliche Lageveränderung der Grundfrequenz hervorgerufen werden könnte.

Um dies zu testen, wurden in zwei Experimenten 14 Sprecher gebeten, in 125 Epochen je vier mal das Wort *beten* zu äußern. Dies Wort wurde gewählt, da es in der Vokalhöhendimension auf der Ebene des betonten Vokals /eː/ zwei unmittelbare Nachbarn im Lexikon aufweist, nämlich *bieten* und *bäten*; hierbei gilt jedoch zu bedenken, dass, während *bieten* und *beten* wahrscheinlich keine selten gebrauchten Wörter bzw. Wortformen sind, die Wortform *bäten* (1. bzw. 3. Person Plural Konjunktiv II Präteritum Aktiv des Verbs *bitten*) zumindest in gesprochener Sprache doch vermutlich eher selten gebraucht wird, d.h. die *lexikalische Frequenz* ist höchstwahrscheinlich gering, was einen *bias* in den Antworten bedingen könnte; auch wurde zwar versucht, keine Dialektsprecher aufzunehmen, aber es ist doch im vorhandenen Sprecherpool mit einer beachtlichen Zahl von Sprechern zu rechnen, deren dialektaler Hintergrund möglicherweise eine größere Toleranz von Veränderungen von /eː/ hin zu offeneren Realisierungen aufweisen könnte (d.h., ihre *Goal Region* könnte größer sein). Diese Frage wurde hier nicht gesondert untersucht, jedoch gab es keine besonderen Auffälligkeiten einer Blockade der Kompensation in eine Richtung, weshalb es m.E. unwahrscheinlich ist, dass ein solcher Effekt aufgetreten sein könnte.

In beiden Experimenten sollten die Perturbationen dazu führen, dass die Versuchspersonen über die auditorische Feedback-Schleife die eigene Stimme hören sollten, die Äußerungen produziert, die in Richtung dieser Nachbarn verschoben zu sein schien, wofür die Sprecher dann kompensieren sollten. Wie das in der Einleitung 3.1 beispielhaft dargestellte *DIVA*-Modell (Guenther et al., 1998; Guenther, Ghosh, Nieto-Castanon & Tourville, 2006; Perkell, 2010) postuliert, steuert zwar eine Feedforward-Kontrolle die Sprachproduktion, aber unter ständiger Nachkontrolle über somatosensorisches und auditorisches Feedback. Wir erwarten also eine Abweichung im auditorischen Feedback im Sinne, dass die dort erwartete sogenannte *Goal Region* nicht erreicht wird. Die Kompensation dient dazu, der auditorischen *Goal Region* wieder nahe zu kommen.

Wie der Literaturüberblick in der Einleitung dieses Kapitels zeigte, sind Kompensationen für Perturbationen des auditorischen Feedbacks im Mittel über alle Versuchspersonen eigentlich nie vollständig, egal ob die Grundfrequenz oder ob Formanten perturbiert worden waren. Dies gilt zumindest dann, wenn nur produzierte Veränderungen im perturbierten Parameter betrachtet werden. Hierfür gibt es mehrere Erklärungsansätze, die – möglicherweise in Kombination – dieses Verhalten erklären könnten:

Erstens ist es möglich, dass – wenn der Spracherwerb erstmal abgeschlossen und die vokalen Organe vollständig ausgewachsen sind – das auditorische Feedback eine untergeordnete Rolle spielt und die Steuerung der artikulatorischen Gesten hauptsächlich über die Feedforward-Kontrolle, eventuell ergänzt durch Überwachung über das somatosensorische Feedback, vonstatten geht. Dies ist vergleichbar für segmentale Eigenschaften schon hypothetisiert worden (Lane et al., 1997), wobei allerdings auch eine ständige Überwachung suprasegmentaler Eigenschaften durch das auditorische Feedback angenommen werden muss; dass durchaus auch auf Perturbationen akustischer Eigenschaften auf segmenteller Ebene recht schnell kompensiert wird, spricht allerdings eher gegen diesen Gedanken (Purcell & Munhall, 2006b).

Zweitens ist ein Konflikt zwischen auditorischem und somatosensorischem Feedback zu verdächtigen, eine vollständige Kompensation zu verhindern (siehe hierzu später mehr).

Auch ein Konflikt zwischen den zwei Formen des auditorischen Feedbacks, nämlich jenem, das die Sprecher über Kopfhörer zugeführt bekommen, und jenem, das durch die Schallleitung durch die Schädelknochen übertragen wird, ist als Grund für Unvollständigkeit der Kompensation denkbar [11].

Drittens ist es vorstellbar, dass nicht allein die Produktion des perturbierten Parameters unter Perturbation verändert wird, um ein Erreichen der auditorischen *Goal Region* zu ermöglichen; dafür sprechen Befunde, dass kompensatorische Antworten auf akustische Perturbationen multidimensional sein und damit mehrere akustische Parameter betreffen können (Villacorta, 2006; Villacorta et al., 2007; Katseff et al., 2010). Zwar postuliert das *DIVA*-Modell eine auditorische *Goal Region*, aber noch weiß man nicht, durch welche akustischen Korrelate diese perzeptuelle Entität definiert ist. Zwar gibt es Untersuchungen, die die Veränderungen der produzierten Werte der ersten beiden Formanten, die durch Perturbation eines der beiden Formanten hervorgerufen werden, und die feststellen, dass nur der perturbierte Formant anders als in der *Baseline* produziert wird (MacDonald et al., 2011); es ist aber vorstellbar, dass in dieser Studie mit den ersten beiden Formanten die falschen Parameter untersucht wurden und sich bei Untersuchung der Grundfrequenz bei *F1*-Perturbationen und des dritten Formanten bei *F2*-Perturbation (um nur zwei Möglichkeiten anzudeuten, die sich natürlich auf die Befunde von (Syrdal & Gopal, 1986) beziehen, dass barkskalierte *F1-f0* bzw. *F3-F2*-Abstände bessere Korrelate für Vokalhöhe bzw. Frontierungsgrad sein könnten) möglicherweise doch multidimensionale Änderungen ergeben hätten. In gewisser Weise ist es also ein Ziel der vorliegenden Untersuchung, einer Definition der auditorischen *Goal Region*, zumindest was Vokalhöhe angeht, ein wenig näher zu kommen. Für die Multidimensionalität sprechen auch zahlreiche Befunde, dass in der Artikulation – und somit auch in der Kompensation für Perturbation – weniger ein Artikulator eine Aufgabe erfüllt (wie die Herstellung der Korrelate für ein phonologisches *feature*), sondern zur Produktion eines Merkmals oft mehrere Artikulatoren involviert sind (siehe hierzu die Beispiele in der Einleitung dieses Kapitels). Dies hat viel damit zu tun, dass natürlich die Artikulatoren mechanisch nicht unabhängig voneinander sind. Andererseits gibt es auch Hinweise, dass sekundär eingesetzte Artikulatoren auch zum sogenannten *feature enhancement* eingesetzt werden, also bestimmte Kontraste möglicherweise aktiv verstärken (siehe z. B. Hoole (2006) oder Hoole und Honda (2011) für die laryngalen Komponenten der Produktion von Vokalen und Plosiven), beziehungsweise bei Perturbation des augenscheinlichsten Artikulators *stattdessen* verstärkt eingesetzt werden (siehe beispielsweise Riordan (1977)).

[11] Die in diesem Experiment benutzten Ohrhörer sind, laut persönlicher Kommunikation zwischen Phil Hoole und Kevin Munhall, möglicherweise nicht die beste Methode, das auditorische Feedback zu unterdrücken, da hierdurch möglicherweise die Weiterleitung des Schalls der eigenen Sprachproduktion durch die Knochenleitung im Schädel (Stenfelt & Goode, 2005) eher verstärkt wird, und zwar gerade im niederfrequenten Bereich einschließlich der Region, in der *F1* sich manifestiert (Pörschmann, 2000); der gleiche Einwand ist auch in Villacorta et al. (2007, Fußnote 2 auf Seite 2318) zu finden, und auch (Purcell & Munhall, 2006b) diskutiert diese Frage, hinweisend auf den Befund, dass die Knochenleitung sehr versuchspersonenabhängig ist (Purcell, Kunov & Cleghorn, 2003), was weitere unnötige Variation in die Ergebnisse von Perturbationsexperimenten einbringt. Als Alternative zu Ohrhörern wird die mit dem auditorischen Feedback gleichzeitig stattzufindende Übertragung von Rauschen vorgeschlagen.

Viertens ist natürlich auch vorstellbar, dass nicht alle Versuchspersonen gleich sensibel für Veränderungen der auditiven *Goal Region* sind und daher weniger vollständig kompensieren. Zumindest wurde schon gezeigt, dass Versuchspersonen, die weniger sensibel einen akustischen Cue zur Unterscheidung von Kategorien nutzen als andere, auch weniger als andere in diesem Parameter bei Perturbation kompensieren (Perkell, Guenther et al., 2004; Perkell, Matthies et al., 2004; Villacorta, 2006; Villacorta et al., 2007). Da Ghosh et al. (2010) und (Brunner et al., 2011) diesen Zusammenhang zwischen der Feinheit der Perzeption und des Ausmaßes der Kompensation für Perturbation auch für die Wahrnehmung auf eher mehrdimensionaler Ebene eines Kontinuums zwischen [s] und [ʃ] zeigen konnten, und Brunner et al. (2011) hierbei auch einen größeren Einsatz von *motor equivalence*, also des sich ergänzenden Gebrauchs mehrerer Artikulatoren bei den feiner Perzibierenden zeigen konnte, ist es vorstellbar, dass auch bei der Betrachtung multidimensionaler Antworten auf Perturbationen (wie in den vorliegenden Experimenten) solche Effekt nach wie vor eine Rolle spielen, und somit grober Perzibierende weniger für Perturbationen kompensieren.

In der Tat fanden wir in beiden Experimenten eine im Mittel über alle Sprecher unvollständige Kompensation im jeweils perturbierten Parameter. Es gibt allerdings auch einige Unterschiede zur Literatur zu berichten.

***F1*-Perturbation** Völlig kompatibel mit den bisherigen Formantperturbationsstudien (siehe Einleitung) ist der Befund der unvollständigen Kompensation, wie er in den letzten beiden Perturbationsphasen zu finden ist. Auch in der vorhandenen Literatur wurde davon berichtet, dass es eine größere Zwischen-Sprecher-Variation geben kann, wie wir sie hier antreffen, also dass es die sogenannten *followers* gibt, die der Perturbation mit der Produktion folgen, und somit den mismatch zwischen dem geplanten Ziel und der erreichten *Goal Region* zusätzlich verstärken, und unter der Kompensierenden ebenfalls starke sprecherspezifische Unterschiede im Ausmaß der Kompensation (die maximal auch „nur“ 55.5% erreicht). Der Mittelwert für Perturbationsstufe 3, in der um 100 mel perturbiert worden war, erreicht mit 24% auch durchaus einen Wert, wie man ihn in der Literatur für Formantkompensation finden kann (MacDonald et al., 2010). Dieser hier gefundene Durchschnittwert ist aber höher als in Studien, die auf lexikalische Nachbarn verzichten und nur ausgehaltene Vokale benutzten, wie z. B. die nur circa 10% Kompensation in Purcell und Munhall (2006b).

Gemittelt über alle Sprecher fanden wir allerdings in den ersten beiden Perturbationsphasen des *F1*-Perturbationsexperiments so gut wie keine Kompensation (gegenüber der *Baseline*). Dies ist so nicht konsistent mit der Literatur. Zwar haben wir keine Rampenphase benutzt, aber MacDonald et al. (2010) hatte gezeigt, dass Rampenphasen, also eine nicht bewußt wahrnehmbare, schleichende Einführung der Perturbation, nicht unbedingt nötig sind, um erfolgreiche Kompensation zu elizitieren; einschränkend muss allerdings erwähnt werden, dass die Stufenhöhe in MacDonald et al. (2010) mit 50 Hz in *F1* doch auch nur etwa die Hälfte dessen betrug, was hier angewendet wurde; seine Perturbationseinführung war also durchaus weniger abrupt. Ein weiterer störender Faktor könnte gewesen sein, dass die *Baseline*-Phase nur aus fünf Epochen bestand, wenn auch durch die pro Epoche

vier Äußerungen von *beten* dieses Wort unperturbiert immerhin 20 mal über (die natürlich immer als künstlich wahrgenommene) auditorische Feedback-Schleife über die Ohrhörer wahrgenommen worden war, bevor zum ersten Mal eine Perturbation einsetzte. Zumindest könnte eine mangelnde Eingewöhnung an die ungewohnte Experimentsituation eventuell nicht nur die (scheinbare) Nicht-Kompensation in den ersten beiden Perturbationsphasen erklären, sondern auch, weshalb nach Epoche 56, also mit Beginn der *RÜCK*-Phase ohne Perturbation, eine recht plötzlich einsetzende überschießende Reaktion eintritt, also eine Verschiebung des produzierten ersten Formanten in Richtung der vorher angewendeten Perturbation. Dies ist so auch nicht konsistent mit den bisherigen Literaturbefunden, wo eher von einer langsamen Deadaptation in vergleichbaren Phasen ohne Perturbation nach erfolgter Perturbation berichtet wurde. Jedenfalls scheinen wir es hier, möglicherweise durch die zu kurze *Baseline*-Phase bedingt, mit einer Art von *Baseline shifting*, der Verschiebung des Bezugspunkts, eventuell ausgelöst durch die generelle Künstlichkeit der Experimentsituation, zu tun zu haben. Wählt man denn auch eine Kombination aus *Baseline-* und *Rückphase* als neue Baseline (hier *BASIS* genannt), ergeben sich auch für die zweite Perturbationsphase Kompensationseffekte in *F1*. Bestätigung findet dieser Umstand auch dadurch, dass auch die Intensität bei den meisten Sprechern unter der Perturbationsrichtung *MINUS* am höchsten, bei der Perturbationsrichtung *PLUS* am niedrigsten ist, mit den Werten für *BASIS* in der Mitte. Wir vermuten hier ein weiteres Korrelat des Öffnungsgrades des Ansatzrohres, eine Art intrinsischer Intensität.

Zur Künstlichkeit der Experimentsituation (und damit zur schweren Eingewöhnung) könnte auch das ausgewählte Material und die Aufgabe beigetragen haben. Zwar wurde darauf geachtet, mit einem zweisilbigen deutschen Wort, dass in der Vokalhöhendimension zwei lexikalische Nachbarn hat, eine möglichst „natürliche“ Auswahl zu treffen, zumindest eine natürlichere als die oft verwendeten ausgehaltenen Vokale, und möglicherweise sogar natürlicher als die ebenso oft eingesetzten /CVC/-Wörter (die in den zumeist englischsprachigen Perturbationsstudien allerdings auch sinnvoller einzusetzen sind als für das Deutsche). Dennoch war die Aufgabe, schnell hintereinander *beten beten beten beten* möglichst ohne eine Art von Satzakzentuierung und auch ohne Deklination zu produzieren, höchst künstlich und bedurfte sicherlich einer längeren Eingewöhnungsphase. Wir können hier dennoch letztendlich keine wirklich überzeugende Erklärung für diesen in den ersten beiden Perturbationsphasen zu beobachtenden Effekt anbieten.

Festzuhalten ist jedoch, dass diese (scheinbare) Nicht-Kompensation nur über alle Sprecher gemittelt auftritt. Betrachtet man einzelne Sprecher, wird man feststellen, dass einige Sprecher in den ersten beiden Perturbationsphasen durchaus kompensiert hatten, aber eine erstaunlich hohe Zahl an Sprechern in diesen beiden ersten Perturbationsphasen das Verhalten von *followers* zeigten, also der Perturbation mit ihrer Produktion folgten. Es gab zwei Gruppen von Sprechern: jene, die zuerst eine Perturbation in Richtung *bieten* zu hören bekamen, und jene, die zunächst mit einer Perturbation in die Gegenrichtung, also nach *bäten* hin, konfrontiert wurden. Da die beschriebenen *followers* in beiden Gruppen auftraten, halten wir einen Effekt der Perturbationsrichtung für unwahrscheinlich. Möglicherweise konnten sich einzelne Sprecher einfach schneller an die ungewohnte Experimentsituation eingewöhnen als andere. Leider führte dieser Effekt die Inkonsistenz bei der

Kompensation in den ersten beiden Perturbationsphasen dazu, dass sich die nachfolgenden Analysen bezüglich des Einsatzes der Grundfrequenz zumeist, wenn auch nicht immer, auf die Perturbationsphasen 3 und 4 beschränken mussten.

Ein weiterer interessanter Umstand ist, dass die produzierten *F1*-Werte in den Perturbationsphasen 3 und 4 in etwa gleich ausfielen. Da sich diese Phasen durch unterschiedliche Perturbationsstärken auszeichnen (mit Verschiebungen um 100 bzw. 200 mel), bedeutet dies, dass die Kompensation für Perturbation in der vierten Perturbationsphase prozentual geringer ausfiel als in der dritten. Dies bedeutet, dass hier ein Plateau in der Produktion erreicht wurde, über das nicht hinausgehend kompensatorisch produziert wurde. Auch MacDonald et al. (2010) stellen ein solches Plateau fest, allerdings bei etwas höheren Perturbationswerten von circa 200 Hz für *F1* (und zwar bei multidimensionaler Perturbation von *F1* und *F2*). Dieser Unterschied könnte durch Sprachspezifika entstanden sein (MacDonald et al. (2010) untersuchte amerikanisches Englisch).

Purcell und Munhall (2006b) beschreiben Kompensationen für *F1*-Perturbationen in phonierter Sprache und stellen fest, dass die Kompensation geringer ausfällt als in geflüsterter Sprache, so wie z. B. in Houde und Jordan (1998, 2002). Schon dies ist ein erster Hinweis darauf, das die Grundfrequenz möglicherweise ebenfalls verändert produziert werden könnte, um ein auditorisches Vokalhöhen-*Goal* zu erreichen – andererseits könnte dieser Effekt einer Verringerung der Kompensation bei phonierter gegenüber geflüsterter Sprache auch einfach auf die bei phonierter Sprache größere Schallleitung durch den Schädelknochen zurückzuführen sein. Genau diese Annahme war der Grund für Houde und Jordan (1998, 2002) gewesen, geflüsterte Sprache zu verwenden.

Anhand des *F1*-Perturbationsexperiments wird die Grundannahme einer gegenläufigen Produktion der Parameter *f0* und *F1* sicher am deutlichsten, denn *F1* ist unbestritten *ein* Vokalhöhenkorrelat, und sicherlich das wichtigste, insbesondere *innerhalb* desselben Sprechers (so wie ja auch in den hier vorliegenden Experimenten die Sprecher die eigene Stimme und somit nur einen Sprecher bezüglich der Vokalhöhe zu bewerten haben). Egal, welches Vokalhöhenkorrelat man als das geeignetste annimmt (*F1* alleine oder eine wie auch immer zu definierende Kombination aus *F1* und *f0*, oder auch eine Kombination von *F2* und *F1*), sie alle inkorporieren den ersten Formanten, was dazu führt, dass eine Perturbation desselben zu einer geänderten Vokalhöhenwahrnehmung führt, und eine unvollständige Kompensation in *F1* für diese Perturbation den in Frage stehenden Vokal als in Perturbationsrichtung verschoben erscheinen lässt. Nehmen wir also nun an, dass eine weitere Kompensation in *F1* geblockt wird, wobei wir in unseren Experimenten keine Aussage darüber machen können, warum dies der Fall ist, da wir andere Feedback-Arten nicht hinreichend kontrollieren konnten, so *müssen* eventuell andere Parameter eingesetzt werden, wenn das auditorische Ziel erreicht werden soll. Denkbar ist z. B. , wie oben erwähnt, dass das somatosensorische Feedback ein mit dem auditorischen Feedback in Konflikt stehendes Signal sendet und deshalb ein Kompromiss zwischen beiden Feedback-Arten gefunden werden muss. So zeigten Larson, Altman, Liu und Hain (2008) eine stärker ausgeprägte Kompensation für *f0*-Perturbation, wenn das somatosensorische Feedback durch Gabe von Anästhetika in den Kehlkopf eingeschränkt und damit das Gewicht afferenter Information mehr zur auditiven Ebene verschoben wurde. Dies macht deutlich, dass Sprecher

nicht allein auf die auditorische Domäne, sondern auch auf die somatosensorische zurückgreifen, und dass deren Feedback offenbar Kompensation beschränken kann; die sprecherspezifisch unterschiedlichen Ausmaße, auf akustische Perturbationen mit Kompensationen zu reagieren, spiegeln vermutlich nur wider, in welchem Ausmaß der jeweilige Sprecher die Informationen aus beiden Domänen gewichtet, und die Unvollständigkeit von Formantkompensationen spiegelt vermutlich die wichtige Rolle, die das somatosensorische Feedback bei der Produktion segmentaler Eigenschaften (wie dem Öffnungsgrad) spielt. Wenn nun also – in Analogie zur *motor equivalence* – andere Parameter eingesetzt werden müssen, so stellt sich die Frage, welche Parameter das sind und in welche Richtung sie verändert produziert werden müssen. In der vorliegenden Studie untersuchen wir den Einfluss der Grundfrequenz auf die Vokalhöhe, und betrachten daher nur diesen Parameter. Wird also beispielsweise *beten* perturbiert, in dem der erste Formant signaltechnisch abgesenkt wird, verschiebt sich – auch bei Kompensation durch Produktion höherer *F1*-Werte, da diese Kompensation unvollständig ist, wie wir gesehen haben – das auditorische Feedback in Richtung *bieten*. Ungeachtet dessen, ob man einen *F1-f0*-Abstand als Vokalhöhenkorrelat annimmt – wobei dieser Abstand nun zu klein für ein /eː/ist –, oder ob man nur die Beeinflussung des Vokalhöhenperzepts bei ambigen Stimuli durch die vokalintrinsische Grundfrequenz (Reinholt Petersen, 1986) als möglichen Grund für eine Involvierung der Grundfrequenz als Grund akzeptiert, ist die Konsequenz für eine erfolgreiche Nutzung der Grundfrequenz dieselbe: nur eine Absenkung der Grundfrequenz kann im Fall einer Perturbation in Richtung *bieten* wieder ein *beten*-Perzept wiederherstellen. Dementsprechend gilt für eine perturbierende Anhebung des ersten Formanten ebenso, dass die Grundfrequenz angehoben werden muss, um das Perzept wieder in Richtung *beten* zu verschieben.

Genau diese Produktionsverschiebung der Grundfrequenz *in Richtung* der Perturbation (oder, in anderen Worten, *in Gegenrichtung* zur kompensatorischen *F1*-Produktionsänderung) stellen wir anhand des ersten Experiments fest. Wir hatten die Methodik aus Villacorta (2006) und Villacorta et al. (2007) übernommen, und erhalten ähnliche Ergebnisse (auch wenn diese, zumindest in Villacorta et al. (2007), eher als Nebenbefund gewertet wurden). Im Gegensatz zu Villacorta (2006) und Villacorta et al. (2007) war es in unseren Daten nicht nötig, für einen generellen Anstieg der Grundfrequenz im Verlauf des Experiments zu normalisieren, da ein solcher Anstieg nicht festgestellt wurde. Über die ersten beiden Perturbatonsphasen eine Aussage zu treffen, ist schwierig, da – wie beschrieben – hier nur von einigen Versuchspersonen in *F1* kompensiert worden war. Zwar ist interessant, dass in der ersten Perturbationsphase im Mittel statt einer kompensatorischen Bewegung von *F1* in Gegenrichtung zur Perturbation *f0* wie hypothetisiert in Richtung der *F1*-Perturbation verschoben produziert wurde; andererseits kann man in Frage stellen, ob es sich hierbei um eine mögliche Vokalhöhenkorrektur mit Mitteln der Grundfrequenz handelt, da dieser Effekt in der zweiten Perturbationsphase, in der stärker perturbiert wurde, nicht mehr auftritt. In den beiden letzten Perturbationsphasen, in denen *F1*-Verschiebungen zur Kompensation genutzt worden waren, zeigt sich jedoch tatsächlich das erwartete Muster, dass *f0* dort steigt, wo *F1* tiefer produziert wird, und umgekehrt. Es gibt aber keine simple negative Korrelation von *f0* und *F1*, die bedeuten würde, dass *f0* umso stärker in eine Richtung verschoben produziert wird, je mehr *F1* in die Gegenrichtung produziert

wird, was man ansatzweise in Villacortas Daten sehen kann, denn in unseren Daten ändert sich im Mittel der produzierte *F1* nicht mit der Perturbationsstärke, wohl aber die Grundfrequenz. Dies, also der Umstand, dass die *F1*-Produktionsänderung schon ab der Perturbationsstärke von 100 mel ihr Maximum erreicht zu haben scheint, lässt die Vermutung aufkommen, dass *stattdessen* die Grundfrequenz zur Vokalhöhenänderung eingesetzt wird, was auch die Daten, über alle Versuchspersonen betrachtet, zu zeigen scheinen. Eine Analyse, die individuell pro Sprecher den *zusätzlichen* Gebrauch der Grundfrequenz unter der Perturbationsstärke 200 mel untersuchte, zeigte aber, dass nur eine leichte Tendenz zu diesem Verhalten gegeben ist, und die Sprecher diesen zusätzlichen Cue zur Kontrastverstärkung nur inkonsistent einsetzen - einige nutzen ihn, einige auch nicht. Die Distanz zwischen produziertem ersten Formanten und Grundfrequenz verändert sich aber, bei aller Zwischen-Sprecher-Variabilität, doch mit der Perturbationsstärke signifikant.

Es stellt sich die Frage, ob *f0* aktiv eingesetzt wird. In der Frage nach der Ursache für vokalintrinsische Grundfrequenz wird oft die Automatik dieses Effekt wegen der mechanischen Kopplung der Zunge mit dem Kehlkopf hervorgehoben (*tongue-pull hypothesis*). Betrachten wir nur eine Perturbationsrichtung. Wird *F1* nach oben perturbiert (in Richtung *bäten* im vorliegenden Fall), wird kompensatorisch *F1* nach unten verschoben produziert (in Richtung *bieten*). Dem Automatismus der mechanischen Kopplung von Zunge und Kehlkopf folgend, müsste also auch ohne aktive Steuerung die Grundfrequenz in diesem Fall steigen – und genau dieses Phänomen beobachten wir in unseren Daten. Von einer aktiven Steuerung der Grundfrequenz zur Verdeutlichung eines Vokalhöhenkontrastes könnte man also nur dann sprechen, wenn *f0* bei der größeren Perturbationsstärke verstärkt eingesetzt werden würde, obwohl sich die *F1*-Produktion nicht nennenswert verändert. Dies ist, wie oben beschrieben, nur für einige Sprecher der Fall. Möglicherweise ist aber gerade dies eine Eigenschaft von Parametern, die zu einem sogenannten *feature enhancement*, also der Verstärkung der Auswirkungen eines anderen Parameters, eingesetzt werden. Ähnliche Beobachtungen machten Hoole und Honda (2011) im Zusammenhang mit Grundfrequenzeffekten bei stimmhaften Konsonanten und vokalintrinsischer Grundfrequenz.

Hoole und Honda (2011) hielten hierzu fest:

> As a guideline for future work, it can be hypothesized that whenever an effect is assumed to be automatic and mechanical then it should be possible to demonstrate that it is fairly constant over speakers. On the other hand, the adoption of enhancement strategies will probably be more variable, reflecting the fact that speakers differ in clarity and, perhaps, in their sensitivity to acoustic differences.

Zusammenfassen ist es also vorstellbar, dass Sprecher, die überhaupt das auditorische Feedback, das ihnen über die Ohrhörer geboten wurde, nutzen, und dementsprechend ihre *F1*-Produktion anpassen, wobei über die mechanische Kopplung der Artikulatoren auch ein gewisser Einfluss auf die Grundfrequenz ausgeübt wird. Wenn nun auditorisches und somatosensorischen Feedback in Konflikt geraten, wird eine weitere Anpassung des ersten Formanten möglicherweise unmöglich gemacht. Die Gruppe der Kompensierenden teilt sich

nun auf in diejenigen, die zur Kontrastverstärkung *f0* aktiv nutzen, um die auditorische *Goal Region* zu erreichen, und jene, die das nicht tun.

Wenn nun aber offenbar *f0* genutzt werden kann und auch genutzt wird, um ein Vokalhöhenperzept zu beeinflussen, beeinflusst dann umgekehrt auch eine Abweichung von der zu produzierenden Grundfrequenz das Vokalhöhenperzept?

***f0*-Perturbation** Im Grundfrequenzperturbationsexperiment gab es ebenso unvollständige Kompensation in *f0* bei erheblicher Zwischen-Sprecher-Variation, was auch konsistent mit der in der Einleitung erwähnten Literatur ist. Die Tatsache, dass von vierzehn Sprechern 3 im gesamten Verlauf des Experiments *followers* blieben, und zwei weitere nahezu vollständig für die *f0*-Perturbation kompensierten, erschwert die weitere Analyse des hier eigentlich interessierenden ersten Formanten. Es ist unklar, wie die Erwartungen bezüglich der wahrgenommen Vokalhöhe bei *followers* sein sollen, und bei den nahezu komplett Kompensierenden ist anzumerken, dass sie – wenn man unsere Annahme bedenkt, dass eine unvollständige Kompensation zu einer Änderung des spektralen *F1-f0*-Abstandes als perzeptives Vokalhöhenkorrelat führen soll – keinen Grund haben sollten, eine Vokalhöhenverschiebung wahrzunehmen; dementsprechend sind für die nahezu komplett Kompensierenden keine kompensatorischen Änderungen des Öffnungsgrades zu erwarten - und somit keine nennenswerten *F1*-Änderungen. Eine Kompensation in Gegenrichtung zur Perturbation, die allerdings unvollständig ausfällt, ist geradezu Bedingung für unsere Grundfrage für einen Einfluss der Grundfrequenz auf die Vokalhöhenperzeption. Somit verbleiben eigentlich nur etwa 9 Sprecher, für die ein kompensatorischer Einsatz des ersten Formanten überhaupt in Frage kommt.

In den ersten drei Perturbationsphasen reagieren die Sprecher im Mittel jedoch schnell auf die Perturbationen, und passen die Produktion auch der Perturbationsstärke an. So entsteht in diesen Phasen ein treppenförmiger Verlauf der produzierten Grundfrequenz, wobei, um im Bilde zu bleiben, die Treppenstufen nur leicht abgerundete Ecken haben, denn die Produktion stabilisisiert sich sehr schnell auf einem Plateau, nachdem innerhalb der ersten circa 5 Epochen noch ein Anstieg zu sehen war. Schon gegen Ende der dritten, hauptsächlich aber in der letzten Perturbationsphase des Grundfrequenzperturbationsexperiments steigt wieder deutlich die Zwischen-Sprecher-Variabilität, d.h. hier werden einige Sprecher mehr zu *followers* bzw. verringern zumindest deutlich das Ausmaß an Kompensation. Dies ist vermutlich schlicht und einfach auf Ermüdung zurückzuführen, da das Formantperturbationexperiment und die Grundfrequenzperturbation direkt hintereinander durchgeführt wurden. Durch die ständige Wiederkehr von *beten*-Äußerungen ist eine Ermüdung nach circa einer halben Stunde sehr wahrscheinlich.

Es ergab sich kein *Baseline-shifting*-Effekt, wie er im *F1*-Perturbationsexperiment gefunden worden war.

Was die Frage betrifft, ob eine unvollständige Kompensation in der Grundfrequenz schon einen Effekt der Vokalhöhenverschiebung hervorrufen kann, für den kompensiert werden muss, sind einige methodische Schwierigkeiten aufgetreten. Zunächst wurden die Daten dieses Experiments genau so analysiert wie die Daten des Formantperturbations-

experiments. Ergebnis schien eine Bewegung des ersten Formanten *in Gegenrichtung* zur Perturbationsrichtung zu sein, was der Hypothese widersprach. Es stellte sich heraus, dass, über alle Sprecher betrachtet, *F1* mit *f0* positiv korreliert zu sein schien. Man könnte dies folgendermaßen deuten: die Veränderung der Grundfrequenz geht mit einer Variation der vertikalen Kehlkopfposition einher, so wie sie K. Honda et al. (1999) gezeigt hat; dort wurde zumindest die Rolle einer Kehlkopfabsenkung als effektiver Einfluss auf eine Grundfrequenzabsenkung beschrieben. Durch diese Lageveränderung des Kehlkopfes variiert die Länge des Ansatzrohrs und kann auf diese Weise die Formantlagen beeinflussen. Chládková et al. (2009) zeigten eine Kovariation der ersten zwei Formanten mit der Grundfrequenz, was (auch) auf die erwähnte Ansatzrohrlängenvariation zurückzuführen sein dürfte. Doch diese Erklärung hat einen Haken: Chládková et al. (2009) zeigte, dass dieser Effekt auch und bei Männern sogar hauptsächlich den zweiten Formanten betraf. In den hier vorliegenden Daten konnte ein solcher Effekt für *F2* aber nicht gefunden werden, auch nicht für *F3*.

Eine sprecherspezifische Analyse der positiven Korrelation zwischen *f0* und *F1* ergab, dass der Effekt über alle Sprecher nur durch besonders ausgeprägte Effekte bei sechs Sprecherinnen zustande kam. Die Analyse der Daten der sechs Sprecherinnen ergab eine erstaunlich genaue Wiedergabe der Frequenzwerte des ersten Formanten durch die Formel $F1 = 2 * f0$. Es schien also sehr wahrscheinlich, dass diese Daten Artefakten des Formantextraktionsalgorithmus entstammten; Fehler solcher Art sind so auch schon für Frauen- und Kinderstimmen beschrieben worden (Iseli et al., 2006) und werden auf den Einfluss der zweiten Harmonischen auf das Messergebnis zurückgeführt. Leider zeigte sich bei Verwendung eines anderen Formantextraktionsalgorithmen keine wesentliche Besserung. Daher wurde beschlossen, trotz der sehr geringen verbleibenden Sprecheranzahl die Analyse ohne diese sechs Sprecherinnen durchzuführen. Leider zeigte sich nun, unter Verwendung der Methodik, die im Test davor angewendet worden war, also grob gesagt durch den Vergleich der Epochen ohne Perturbation mit denen mit Perturbation, kein Effekt in *F1*. Wir haben in diesem Experiment *f0*-Perturbationen im Bereich von ± 2 Halbtönen vorgenommen. Dies entspricht – gegeben, ein Sprecher kompensiert überhaupt nicht – einem Umfang von 5 Halbtönen an Variation. Für diese Perturbation wird im Mittel aber um 26% kompensiert (miteingerechnet die sogenannten *followers*), d.h. der Umfang der Variation um 5 Halbtöne wird zusätzlich noch im auditorischen Feedback reduziert. Es stellt sich die Frage, ob diese Variationsbreite groß genug ist, um eine Kompensation der Vokalhöhe zu erwarten. Nimmt man jedenfalls den Umfang, den jeder Sprecher hörte (für jeden Sprecher, gemäß seiner Kompensationsfähigkeit, unterschiedlich) als Ausgangspunkt, und untersucht nicht einfach getrennt danach, ob es Perturbation gab oder nicht, sondern nach Perturbationsrichtung, ergibt sich ein leichter Effekt auf *F1* in die hypothetisierte Richtung, also *in Gegenrichtung* zur Bewegung der Grundfrequenz. Doch nicht nur die Extreme unterscheiden sich: auch zwischen der *BASIS*-Stufe und der *MINUS*-Perturbation ergab sich ein Effekt, jedoch nicht zwischen der *BASIS* und der *PLUS*-Richtung. Eine Analyse der Intensität kann hier eventuell als Erklärung herhalten, denn ungeachtet von der Perturbationsrichtung stieg die Intensität unter Perturbation gegenüber den Epochen ohne. Möglicherweise wird also unter Perturbation mit der vokalen Intensität der Grad der Kieferöffnung, und damit *F1* erhöht,

was, wie bereits in der Einleitung erwähnt wurde, ein konsistent zu findendes Muster für Erhöhung vokaler Intensität ist. Dennoch weist auch der Unterschied in *F1* zwischen *BASIS* und der *PLUS*-Richtung die erwartete Richtung auf, wenn auch eben nur tendenziell.

Interessant ist auch der Blick auf die Formanten 2 und 3. Tatsächlich variieren beide Parameter mit der kompensatorischen Grundfrequenzbewegung, wenn die Änderungen in *F2* auch nur tendenziell waren, aber bei *F3* als signifikant bewertet wurden; dies spricht sehr für eine Kovariation von Grundfrequenz und Kehlkopfhöhe, mit den beschriebenen Auswirkungen auf die Formanten durch Ansatzröhrlöngenvariation.

Umso deutlicher ist dann allerdings der in gegenläufige Richtung weisende Befund für *F1*, da dieser Parameter ja dann auch von dieser Ansatzrohrlängenvariation beeinflusst sein müsste, und zwar in Gegenrichtung zu der, die beobachtet wurde.

In der Tat verhält es sich denn auch so, dass eine weitere Analyse offenbarte, dass nicht alle Sprecher das hier beschriebene Muster einer gegenläufigen Bewegung von *f0*- und *F1*-Produktion zeigen. Ein Grund ist sicherlich der bereits erwähnte, dass Sprecher, je mehr sie prozentual kompensieren, umso weniger Grund haben sollten, überhaupt eine Vokalhöhenbeeinflussung wahrzunehmen, und dementsprechend auch nicht kompensieren müssen. In der Tat zeigt Abbildung 3.36, dass derjenige Sprecher, der am meisten kompensierte, fast keine Änderung in *F1* vorzuweisen hat (allerdings eine generelle Erhöhung des *F1*-Wertes gegenüber der *Baseline*, was mit der oben beschriebenen Erhöhung der Intensität in perturbierten Phasen erklärt werden könnte). In der Tat zeigt die gleiche Abbildung auf Seite 147, dass es Versuchspersonen gibt, die wenig kompensieren, aber viel *F1* variieren – aber es gibt auch Gegenbeispiele, mit viel *F1*-Variation in die „falsche“ Richtung. Der einzige *follower* unter den acht übriggebliebenen Versuchspersonen zeigt allerdings, in dem er auch *F1 in Richtung* der Perturbation verschiebt, ein eigentlich zu erwartendes Verhalten, denn, da er der Perturbation folgte, erweiterte er ja die die Vokalhöhe möglicherweise beeinflussende Variation der Grundfrequenz, und hatte somit eigentlich noch mehr Grund als die anderen Versuchspersonen, deswegen auch *F1* in die gleiche Richtung zu verschieben.

Wir halten also fest, dass auch unter Grundfrequenzperturbation eine Wahrnehmungverschiebung in Richtung einer anderen Vokalhöhe möglich ist, für die kompensiert werden kann. Dieser Effekt, der gegenläufig zur Grundfrequenzveränderung verläuft, ist vorhanden, obwohl eine dieser Bewegung entgegenstehende Variationsquelle in Form einer Vokaltraktlängenvariation und damit eine *in Richtung* der Grundfrequenzproduktion zu erwartenden Formantvariation wahrscheinlich ist.

Es ist durchaus beachtenswert, dass ein zur *f0*-Bewegung gegenläufiger Trend für *F1* festgestellt werden kann. Wie wir in Bezug auf die gegenläufige Bewegung beider Parameter im *F1*-Perturbationsexperiment diskutiert haben, ist es durchaus vorstellbar, dass die *f0*-Bewegung zumindest teilweise durch mechanische Kopplung der Artikulatoren entstanden sein kann, und es gibt nur für eine Teilmenge der Sprecher Hinweise für einen aktiven Einsatz der Grundfrequenz. Im Falle des *f0*-Perturbationsexperiments jedoch ist mit einer solchen Kopplung, also einem Automatismus, m.E. nicht zu rechnen; eine Kopplung ist eher in gegenläufige Richtung zu erwarten, nämlich durch Variation der vertikalen Kehlkopflage, und für eine solchen Kopplung würden auch die Ergebnisse für *F2* und v.a. *F3* sprechen. Alle Sprecher, die hier *F1* in Gegenrichtung verschoben haben, sollten dies also aktiv

gesteuert haben. Daraus folgt wiederum, dass zumindest diese Sprecher eine Verschiebung der Vokalhöhe wahrgenommen haben müssen.

Dass also ein Vokalhöheneinfluss damit auch für *f0*-Perturbationsexperimente gefunden werden kann, ist durchaus erstaunlich, denn die beiden hier präsentierten Experimente weisen durchaus die Vokalperzeption betreffend einige deutliche Unterschiede auf. Hiermit ist nicht nur die geringere Rolle von *f0* gegenüber *F1* als Vokalhöhenkorrelat gemeint, oder dass für *f0*-Perturbation mit mehr Kompensation gerechnet werden muss, was die Aussichten auf eine Beeinflussung des Vokalhöhenperzepts weiter verringert, sondern jene in den Methodenteilen beider Experimente beschriebenen Unterschiede in der Art der Perturbation – *lokal* vs. *global*.

Ein weiterer, entscheidender Unterschied zwischen den Experimenten ist sicherlich nämlich auch, dass die Perturbation im *F1*-Experiment nur das /eː/ aus *beten beten beten beten* betraf, im *f0*-Experiment jedoch immer applizierte, also bei jeder Äußerung, die die Versuchsperson machte, was z. B. auch Zwischenfragen usw. betraf. Es wurde durch Befragen der Versuchspersonen nach den Aufnahmen auch deutlich, dass diesen bei dem *f0*-Perturbationsexperiment die Perturbierung ihrer Sprache offenbar wesentlich bewusster war als im *F1*-Perturbationsexperiment. Mehr Sorgen macht uns in diesem Zusammenhang aber ein anderer Punkt. Alle spektralen Eingenschaften blieben im Formantperturbationsexperiment gleich, mit Ausnahme der des Vollvokals in /beːtn/. Es gibt hier keine Quelle, die eine Rekalibrierung im Sinne einer vokalextrinsischen Normalisierung wie in Ladefoged und Broadbent (1957) nötig gemacht hätte. Wie gesagt applizierte der *f0*-Shift aber immer, die Versuchspersonen hörten sich also immer zu tief oder zu hoch. Zwar gab es nur einen Vollvokal in den Äußerungen, dennoch ist aber die Frage zu stellen, ob diese globale Verschiebung nicht dazu hätte führen sollen, dass sie sich schnell an diese neue Stimme gewöhnen und für deren Eigenheiten normalisieren. Dies ist auch für die Anwendbarkeit der hier gemachten Aussagen auf die Erklärung der Kovariation von *f0* und *F1* mit dem Alter interessant, da ja nicht angenommen werden kann, dass die altersbedingten Grundfrequenzänderungen nur lokal auftreten. Warum also gewöhnt sich nicht einfach ein Sprecher an diese „neue“ Stimme, ohne *F1* variieren zu müssen? Dieser Frage müssen wir nachgehen, indem wir im folgenden Kapitel überprüfen, ob Versuchspersonen auf lokale und globale Grundfrequenzänderungen reagieren und inwieweit deren Einfluss das Vokalhöhenperzept der Hörer beeinflusst.

Kapitel 4

Beitrag der Grundfrequenz zur Vokalklassifikation

4.1 Einleitung

Bislang haben wir in dieser Arbeit festgestellt, dass bei alternden Erwachsenen *f0* und *F1* kovariieren, und die uns als wahrscheinlichste These erscheinende Behauptung aufgestellt, dass sie dies tun, da eine sich altersbedingt ändernde Grundfrequenz auf die Vokalhöhenperzeption auswirken könnte. Diese natürliche Perturbation führe dann zur Notwendigkeit, durch Variation des Öffnungsgrades des Kiefers für diese Verschiebung der Vokalhöhenperzeption zu kompensieren. Desweiteren stellten wir fest, dass für Vokalhöhenperturbationen über quasi-Echtzeit-Verschiebung des ersten Formanten im auditiven *feedback* des Sprechers auch mit einer Variation der Grundfrequenz, und nicht nur mit einer Variation des ersten Formanten, kompensiert wird. Zumindest manche Sprecher scheinen hierbei die Grundfrequenz auch unabhängig von *F1* zu nutzen, was dafür spricht, dass sie tatsächlich *f0* als vokalhöhenbeeinflussendes Merkmal nutzen. Auch in einem Grundfrequenzperturbationsexperiment konnten wir Hinweise darauf finden, dass eine Verschiebung der Grundfrequenz die Vokalhöhe beeinflussen kann, wofür dann durch Variation des ersten Formanten kompensiert wird.

Alle diese Befunde sprechen also dafür, dass Sprecher/Hörer eine wie auch immer geartete Kombination von *F1* und *f0* als Vokalhöhenkorrelat nutzen, wobei man sich unter dieser Kombination so etwas wie den *F1-f0*-Abstand in Bark, den mehrere Forscher als entscheidendes Maß annahmen/annehmen (so etwa Traunmüller (1981)) vorstellen kann. In all den oben genannten bisherigen Ergebnissen ist also die *intrinsische* Information als Haupteinflussfaktor zu bestimmen. Wir wollen in diesem Kapitel ein wenig Licht werfen auf die Frage, inwieweit diese *intrinsische* Information in fließender Rede überhaupt genutzt wird, und wie sich *extrinsische* Faktoren auf die Nutzung der *intrinsischen* für die Vokalwahrnehmung auswirken.

Wie in der Einleitung dieser Dissertation schon kurz angedeutet, zeigen einige (auch schon sehr frühe) Forschungen, dass nicht alleine intrinsische Faktoren zur Vokalidentifika-

tion durch menschliche Hörer beitragen (z. B. Ladefoged und Broadbent (1957), Ainsworth (1975), oder Nearey (1989)). Um nur die früheste und wahrscheinlich bekannteste dieser Arbeiten zu rekapitulieren: Ladefoged und Broadbent (1957) zeigten, dass ein synthetisiertes Zielwort, in dem alles unverändert bleibt, also z. B. auch die *intrinsische* Information, die den Vokal formt, anders wahrgenommen und damit der Vokal anders kategorisiert wird, je nachdem, welche Lage die Formanten in den Vokalen in dem synthetisierten Satz, in den die Zielworte eingebettet waren, aufwiesen. Eine Variation der Formanten (*F1* und/oder *F2*) in den Vokalen des Trägersatzes veränderte also die Perzeption des Vokals im Zielwort, was so gedeutet werden kann, dass die Vokale im Trägersatz Beispiele für Vokale eines Sprechers sind, und der Hörer entnimmt aus diesen die Information, wie der Vokalraum abzugrenzen ist. Die Vokale des Zielworts werden dann in Relation zu diesem Vokalraum eingeordnet.

In gewisser Weise werden in diesem Versuchsaufbau also die Hörer mit unterschiedlichen Vokalräumen unterschiedlicher „Sprecher“ konfrontiert, wofür sie normalisieren. Johnson konnte zeigen, dass solche Unterschiede in wahrgenommener Sprecheridentität auch durch die Lage (Johnson, 1989a) und die Spannweite (Johnson, 1990) der Grundfrequenz beeinflusst werden kann, und dass diese wahrgenommenen Sprecherunterschiede zu unterschiedlichen Kategorisierungen von akustisch eigentlich identischen Stimuli führt.

Wir wollen hier einem ähnlichen Ansatz folgen, und *intrinsische* Effekte der Grundfrequenz auf die Vokalhöhenwahrnehmung unter unterschiedlichen *extrinsischen* Bedingungen untersuchen. Da wir, in Gegensatz zu den Vorgängerstudien, keine synthetischen Stimuli verwenden, sondern Resynthesen der Sprache eines Sprechers, und wir die Grundfrequenzvariation auf dessen Spannweite der von ihm genutzten intrinsischen Grundfrequenz beschränken wollen, nehmen wir *nicht* an, dass die Versuchspersonen die Wahrnehmung unterschiedlicher Sprecher haben werden - ganz so, wie alternde Sprecher keine „fremde“ Stimme wahrnehmen werden, wenn sie die eigene, sich langsam verändernde Stimme hören. Zusätzlich werden wir durch den einleitenden Text zum Experiment ausdrücklich darauf hinweisen, dass alle gehörten Äußerungen von *einem* Sprecher stammen.

Wir wollen in Trägersätze eingebettete Stimuli nutzen, bei denen die Grundfrequenz *global* variiert wird. Nimmt man an, dass der *F1-f0*-Cue eine Rolle für die Vokalhöhendifferenzierung im hier verwendeten Deutschen ist, so müsste also eine *globale* Grundfrequenzverschiebung das Vokalhöhenperzept in *allen* in der Äußerung vorkommenden Vokalen beeinflussen. Da wir die natürliche Variation an Grundfrequenz in der gegebenen Sprecherstimme (für den gegebenen Kontext) nicht überschreiten wollen, brauchen wir als Zielstimulus einen ambigen Stimulus, um auch eine durch eine so geringe Variation der Grundfrequenz hervorgerufene Kategorisierung in die eine oder andere Richtung zu elizitieren (vergleiche hierzu Reinholt Petersen (1986), der die Effekte geringer Grundfrequenzvariation auf die Kategorisierung auf durch Formantmanipulation erzeugte Vokalkontinua zeigte; die Effekte waren aber gering, und betrafen nur den ohnehin spektral ambigen Bereich des Kontinuums). Im Falle der *globalen* Grundfrequenzverschiebung erwarten wir folgerichtig auch keinen oder einen nur gering ausgeprägten Effekt der Grundfrequenz auf die Vokalhöhenperzeption, da – im Vergleich zu den anderen Vokalen des Trägersatzes – der Stimulus unverändert ambig bleiben sollte; wir erwarten also, dass die Hörer Schwierigkeiten haben werden, den ambigen Stimulus überhaupt zuordnen zu können. Sollten *intrinsische* Pa-

rameter nicht vollständig von den *extrinsischen* Eigenschaften geblockt werden und zur Perzeption herangezogen werden können, sollte ein leichter Effekt einer eher kontinuierlichen Vokalwahrnehmung für den Zielstimulus zu finden sein.

Dem gegenübergestellt wollen wir aber auch eine *lokale* Grundfrequenzverschiebung testen, indem wir im Trägersatz die Grundfrequenz (und damit die Vokalhöhenwahrnehmung) nicht manipulieren, sondern nur in der ambigen Silbe des Zielworts. Hierfür erwarten wir einen deutlichen Effekt des *intrinsischen* Merkmals *f0*, und somit eine Wahrnehmung, die eher kategorialer Natur sein dürfte. Wir nehmen dies an, obschon uns bewußt ist, dass die lokale Grundfrequenzvariation auch den Intonationsverlauf beeinflussen wird. Da in den letzten Jahren demonstriert wurde, dass zumindest Sprecher/Hörer von Sprachen mit Vokalsystemen mit reicher Vokalhöhendifferenzierung wohl zwischen *intrinsischen* Effekten auf die Grundfrequenz und der intonatorischen Funktion der Grundfrequenz unterscheiden können (*intrinsic pitch*, vgl. die Befunde von Chuang und Wang (1978); Stoll (1984); Fowler und Brown (1997); Niebuhr (2004); Pape (2005); Pape, Mooshammer, Fuchs und Hoole (2005); Pape und Mooshammer (2006b, 2008)), nehmen wir an, dass dieser für die Vokalhöhendifferenzierung hinderliche Effekt der Intonation schwächer sein wird als die generelle Vokalhöhenverschiebung, die wir für die *globale* Grundfrequenzverschiebung annehmen. Diese Annahme beruht darauf, dass der Stimulus spektral ambig sein wird, und wir deshalb vermuten, dass die Gewichtung der Grundfrequenzinformation zwischen den *intrinsischen*, vokalhöhendefinierenden Merkmalen und den intonatorischen Merkmalen durch die Hörer in einem solchen Fall der Uneindeutigkeit des Vokaltokens in Richtung der *intrinsischen*, vokalhöhendefinierenden Information verschoben werden wird.

Insbesondere für das Deutsche konnte gezeigt werden, dass sich [$\pm tense$]-Paare wohl in der letztendlich erreichten Zungenposition, nicht aber in der intrinsischen Grundfrequenz unterscheiden, was ein Hinweis darauf ist, dass in den ungespannten Vokalen *f0* aktiv angehoben wird, was auch durch physiologische Befunde gestützt wird Fischer-Jørgensen (1990); Mooshammer et al. (2001); Hoole et al. (2004); Pape und Mooshammer (2004, 2006a); Hoole (2006); Hoole und Honda (2011). Daraus ist zu schließen, dass Hörer des Deutschen eventuell dem Grundfrequenz-Cue für Vokalhöhe in ungespannten Vokalen mehr Gewicht zuordnen könnten als in gespannten Vokalen, da in den gespannten Vokalen die intrinsische Grundfrequenz weniger aktiv gesteuert zu sein scheint als in den ungespannten. Auch dies wollen wir testen, indem wir einem gespannten Zielwortpaar (*bieten-beten*) ein ungespanntes im gleichen Kontext (*bitten-betten*) gegenüberstellen wollen.

Fassen wir also zusammen:

- Die Grundfrequenzvariation wird grundsätzlich einen Einfluss auf die Vokalperzeption haben
- Ein lokales Grundfrequenzkontinuum resultiert in einer steileren Antwortkurve als ein globales
- Die Antwortkurve für [$-tense$]-Stimuli wird steiler sein als die für [$+tense$]-Stimuli

4.1.1 Kategorialität der Perzeption

Abbildung 4.1 zeigt stilisierte, idealtypische Identifikationskurven zwischen zwei Kategorien, x und y. Nehmen wir an, Stimulus 1 repräsentiert einen typischen Vertreter von x, und Stimulus 11 einen typischen Vertreter von y. Die schwarze Identifikationskurve und die schwarzen Datenpunkte zeigen das Extrembeispiel für eine kategoriale Wahrnehmung: Stimuli 1 bis 5 werden zu jeweils 100% als x, Stimuli 6 bis 11 zu 100% als y wahrgenommen. Es existiert eine Kategoriengrenze, die als Umkipppunkt der Identifikationskurve berechenbar ist (hier 5.5).

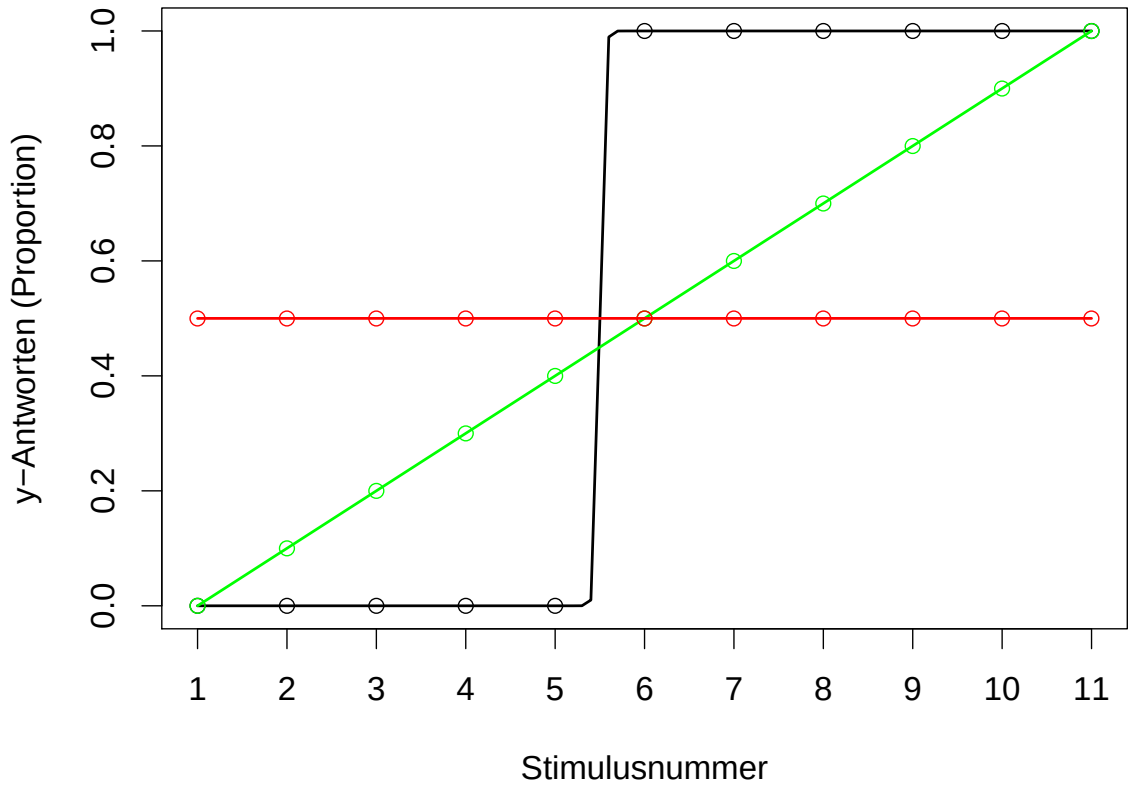

Abbildung 4.1: *Stilisierte, idealtypische Identifikationskurven zwischen zwei Kategorien,* x *und* y. *Siehe Text für Details.*

Die roten Punkte und die dazugehörige Gerade bedeuten das genaue Gegenteil, nämlich dass alle Stimuli nicht zugeordnet werden können. Die gleiche Indentifikations-„Kurve“ könnte so auch entstehen, wenn nicht alle Datenpunkte bei 0.5 liegen, aber um den Mittelwert 0.5 herum. Ebenso flach könnte die „Kurve“ sein, wenn es einen sogenannten *bias* in

Richtung von x oder y geben würde, aber keinen durch das Kontinuum gesteuerten Unterschied in der Wahrnehmung. In all diesen Fällen gibt es keine Kategoriengrenze. Diese Form ist eher theoretischer Natur, da wir als Prämisse davon ausgegangen waren, dass Stimulus 1 und Stimulus 11 eindeutige Vertreter je einer von zwei unterschiedlichen Kategorien sind. Dennoch gibt es Identifikationskurven, die sich dieser Form zumindest annähern.

In grün wird eine dritte Möglichkeit in einer besonderen Ausprägung gezeigt: zwar werden die Endpunkte zu 100% den Kategorien x bzw. y zugeordnet, die Wahrnehmung entlang des Kontinuums ist aber nicht mehr kategorial, sondern kontinuierlich, und die Kurve im Extremfall linear; dies bedeutet, dass die Wahrnehmung sehr fein durch die Details bestimmt ist, und das Kontinuum nicht in zwei Kategorien eingeteilt wird. Auch eine solche „Kurve“ kann, wie übrigens natürlich auch jene der kategorialen Wahrnehmung (schwarz), durch einen bias in die eine oder andere Richtung verschoben sein. Je nachdem, ob die jeweiligen Enden zu den beiden unterschiedlichen Kategorien zugeordnet werden können oder nicht, kann ein Umkipppunkt berechnet werden oder nicht, d.h. eine Art von Kategoriengrenze kann durchaus vorhanden sein, wenn diese auch nur angibt, dass ab diesem Punkt jeweils die Mehrheit der Antworten in die eine oder andere Richtung weisen. Bei diesem Typ ist ohnehin die Steigung der Kurve entscheidender; je flacher diese ist, desto weniger kann zwischen den Endpunkten unterschieden werden - bis hin zur Steigung 0, die bedeutet, dass die Versuchspersonen überhaupt nicht in der Lage sind, zwischen den zwei Kategorien zu unterscheiden - womit wir wieder bei der roten „Kurve“ wären.

4.2 Methode

4.2.1 Vorbereitung der Kontinua

Ziel des Perzeptionsexperiments ist es, den Einfluss von Grundfrequenzvariation auf die Wahrnehmung von Vokalkategorien bei spektral ambigen Stimuli zu ermitteln. Zum Zwecke der Kontinuumserstellung für diesen Zweck sollten folgende Bedingungen eingehalten werden. Erstens sollten die Stimuli so natürlich als möglich sein; aus diesem Grunde wurden zunächst Aufnahmen eines Sprechers des Standarddeutschen als Grundlage für Resynthesen benutzt. Zweitens sollten die Stimuli spektral uneindeutig sein, d. h. ambig bezüglich der Zuordnung zu Vokalkategorien, um allein den Einfluss der intrinsischen Grundfrequenz ermitteln zu können; um dies zu erfüllen, wurden zunächst Kontinua erzeugt, die bezüglich der intrinsischen Grundfrequenz normalisiert wurden, indem stets die gleiche Grundfrequenz bei der Resynthese benutzt wurde, während die Spektra gemorpht wurden. Die so entstehenden Kontinua sollten zwar eindeutige, unterschiedlichen Vokalen zuzuordnende Endpunkte haben, jedoch sollten diese Perzepte nicht durch Einflüsse der Grundfrequenz verstärkt werden. Die spektralen Eigenschaften derjenigen Tokens aus diesen Kontinua, die sich als perzeptiv ambig erweisen würden, wären dann die Grundlage für jene Kontinua, die ausschließlich den Einfluss der Grundfrequenz auf die Vokalperzeption ermitteln sollen.

4.2.2 Sprachaufnahmen

Ein 34-jähriger, in Schleswig-Holstein aufgewachsener männlicher Sprecher der norddeutschen Variante des Standarddeutschen wurde gebeten, je 5 Wiederholungen von Sätzen der Form *Ich habe ZIELWORT gesagt* zu sprechen. Das Korpus enthielt neben einigen Füllwörtern Tokens von *bieten*, *beten*, *bitten* und *betten*. Siehe Tabelle B.1 für eine vollständige Übersicht über das aufgenommene Sprachmaterial.

Die Aufnahmen fanden im Tonstudio des Instituts für Phonetik und Sprachverarbeitung der Ludwig-Maximilians-Universität München in einer schallgeschützten Sprecherkabine statt. Es wurden zwei Kanäle benutzt: auf einem Kanal wurde das Signal des Kondensatormikrophons ™Neumann TLM 103, auf dem anderen Kanal das Signal eines Nackenbügelmikrophons ™Beyerdynamic Opus 54 aufgenommen. Für die Erstellung der späteren Kontinua wurde der Headsetkanal benutzt. Die Aufnahmen wurden mit *MAuS* automatisch segmentiert und etikettiert, die Segmentgrenzen wurden händisch nachkorrigiert. Die TextGrids aus Praat wurdem mit *praat2emu* in Emu-lesbare files konvertiert und in einer Emu-Datenbank zusammengefasst. Mittels *forest* wurden Formanten berechnet, und zwar unter Benutzung eines 25 ms langen Blackman-Analysefensters und einer Fensterverschiebung von 5 ms; als Nominalfrequenz für F1 wurde der Standard für Männer, also 500 Hz, gewählt. Mit dem in VoiceSauce implementierten STRAIGHT-Grundfrequenz-Tracker wurden die f0-Werte ermittelt. Dies geschah, wie es in diesem Algorithmus üblich ist, jede Millisekunde. Emu-R wurde eingesetzt, um die Daten auszulesen. Die Abbildung 4.2 zeigt die intrinsischen Grundfequenzen für die Vokale /ɑː ɛ eː ɛː ɪ iː oː uː ʏ/ aus dem /ʔɪçhɑːbəbVtᵊngəzɑːkt/-Kontext, gemessen in den mittleren 20% der Vokaldauern, also um den jeweiligen zeitlichen Mittelpunkt der Vokale herum, und als Medianwerte gemittelt. Deutlich ist der Anstieg der Grundfrequenz mit zunehmender Vokalhöhe zu sehen, solange man nur je einen Gespanntheitsgrad betrachtet. Dies entspricht sicherlich den Erwartungen, die man für einen Sprecher des Standarddeutschen haben kann, wenn man die Literaturbefunde bezüglich intrinsischer Grundfrequenz betrachtet. Interessanter ist jedoch die Tatsache, dass der vorliegende Sprecher in den gespannt/ungespannt-Paarungen des /bVt/-Kontextes (also *bieten* vs. *bitten* und *beten* vs. *betten*) stets bei den ungespannten Vokalen eine höhere Grundfrequenz als bei den gespannten produziert (bei beiden Paaren beträgt der Unterschied gemittelt in etwa einen halben Halbton, nämlich 0,45 Halbtöne bei [ɛ-eː] und 0,62 Halbtöne bei [ɪ-iː]). Die tiefsten IF0-Werte sind erwartungsgemäß bei [ɑː] zu finden (im Mittel 90,6 Hz), die höchsten jedoch bei [ɪ] (112.4 Hz), was einem Unterschied von 3,73 Halbtönen entspricht. Vom maximalen *f0*-Wert bei [ɪ] (119,2 Hz) bis zum Minimum bei [ɑː] (88,4 Hz) sind es 5.2 Halbtöne.

4.2.3 Ermittlung spektral uneindeutiger Vokale zwischen [iː] und [eː] bzw. [ɪ] und [ɛ].

Als Vorbereitung für die Perzeptionsexperimente zum Einfluss der Grundfrequenz auf die Vokalwahrnehmung mussten zunächst perzeptiv ambige Vokaltokens erzeugt werden. Hierzu wurden Vokalkontinua (von *bieten* nach *beten* und von *bitten* nach *betten*) ohne Grund-

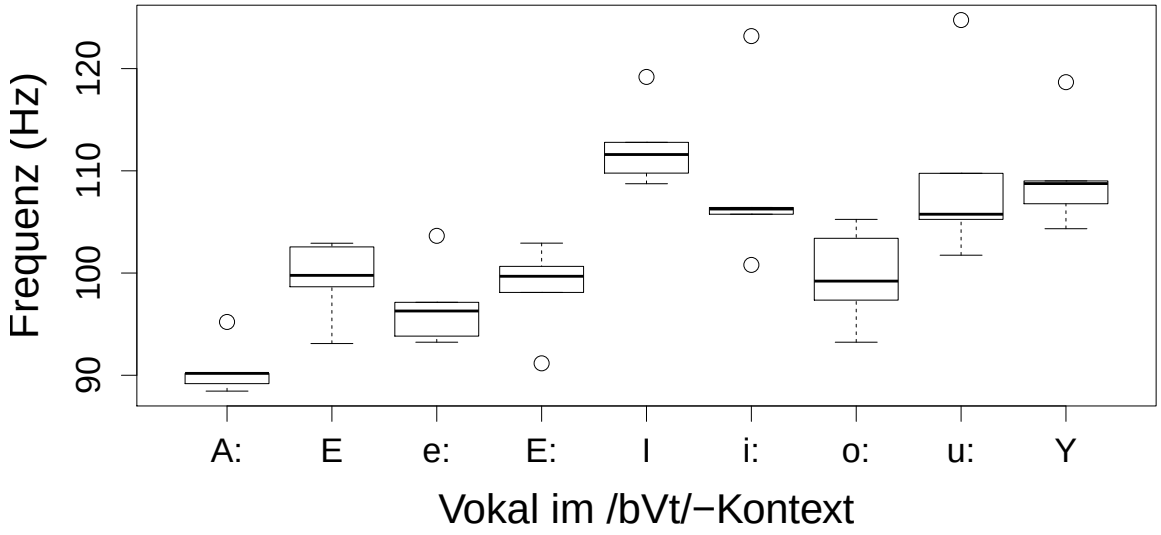

Abbildung 4.2: *Intrinsische Grundfrequenz, ermittelt als Medianwerte in den mittleren 20% der Vokaldauer, der Vokale* /ɑː ɛ eː ɛː ɪ iː oː uː ʏ/ *(in der Abbildung SAMPA-kodiert) im* b*V*t*-Kontext, also für die in den Trägersatz* Ich habe ZIELWORT gesagt *eingebetteten, jeweils 5 Tokens von* baten, betten, beten, bäten, bitten, bieten, boten, booten (/buːt$^{(ə)}$n/), Bütten *für den Sprecher des Standarddeutschen, dessen resynthetisierten Aufnahmen von* beten, bieten *und* betten, bitten *für die Perzeptionsexperimente dieses Kapitels herangezogen wurden.*

frequenzunterschiede erzeugt, in denen ausschließlich spektrale Eigenschaften das Vokalperzept beeinflussen.

4.2.4 Auswahl von Tokens aus der Sprachdatenbank

Zunächst wurden aus der im vorigen Unterkapitel angesprochenen kleinen Sprachdatenbank geeignete Tokens ausgesucht. Es handelte sich um ein *bieten*-token mit einer Grundfrequenz von 106 Hz, einem ersten Formanten von 259 Hz und einem zweiten Formanten von 2101 Hz, ein *beten*-token mit 103 Hz als f0, 299 Hz in F1 und 2186 Hz in F2, sowie ein *bitten*-token mit einer Grundfrequenz von 112 Hz, einem F1 von 360 Hz und einem zweiten Formanten von 1659 Hz und ein *betten*-token mit einer f0 von 102 Hz, dem F1-Wert von 486 Hz und dem F2-Wert von 1577 Hz, wobei die genannten Formantwerte als Medianwerte der mittleren 20% der Vokaldauern ermittelt wurden. Die Auswahl der Tokens richtete sich nach den Werten von F1: für [iː ɪ] wurde jene Tokens mit den höchsten F1-Werten ausgesucht, für [eː ɛ] jene mit den niedrigsten F1-Werten, so dass F1 als Vokalhöhenkorrelat möglichst wenig gestreut in die Kontinua für die Vortests einfloss, so dass möglichst feine Unterschiede

zwischen den zu erzeugenden Stimuli zu erwarten waren. Der Autor hat die ausgesuchten Wörter ohrenphonetisch daraufhin überprüft, dass sie eindeutig der jeweiligen Kategorie zugeordnet werden konnten.

4.2.5 Morphing in TANDEM-STRAIGHT

Die ausgesuchten Tokens wurden in TANDEM-STRAIGHT (Kawahara et al., 1999; Kawahara & Irino, 2005; Kawahara et al., 2008; Kawahara, Nisimura et al., 2009; Kawahara, Takahashi, Morise & Banno, 2009), einer Art von Mehrkanal-Vocoder, analysiert (d. h. auch in Quell- und Filtersignal zerlegt, wobei das Quellsignal nicht nur durch eine Grundfrequenzanalyse, sondern auch durch ein Aperiodizitätsspektrum für stimmlose Signalanteile repräsentiert werden kann); die Analysen wurden in Form von *m*-Dateien, einem Matlab-Format, gespeichert. STRAIGHT erlaubt es, aus diesen *m*-files Resynthesen von erstaunlich natürlicher Qualität zu erzeugen, wobei auch zwischen zwei beliebigen Signalen gemorpht werden kann. Hierbei können - anders als bei anderen Morphing-Methoden - mehrere signaltechnische Parameter gleichzeitig oder auch getrennt voneinander manipuliert bzw. gemorpht werden. Es handelt sich hierbei u. a. um temporale Aspekte, den Intensitätsverlauf, den Verlauf der Grundfrequenz und um die Einhüllende des Spektrums (Manipulationen einzelner Formanten sind nur mit Zusatzsoftware möglich; man beachte jedoch die Möglichkeit, Ankerpunkte zu setzen, s. u.). Für eine erfolgreiche Resynthese gemorphter Signale muss eine Art Segmentation über sogenannte Ankerpunkte angefertigt werden, die die akustischen Landmarken, die korrespondierend in beiden Signalen vorhanden sind, abdecken sollte; diese sogenannten Ankerpunkte sind auf zwei Ebenen zu setzen: einerseits auf temporaler Ebene, um beispielsweise On- und Offset eines Vokoiden zu setzen, andererseits auf spektraler Ebene, in der Hauptsache um Formanten zu markieren.

Um also Kontinua zwischen *bieten* und *beten* sowie zwischen *bitten* und *betten* anzufertigen, wurden zunächst die Endpunkte erzeugt, in denen alle akustischen Eigenschaften mit Ausnahme der Grundfrequenz aus den Originalaufnahmen übernommen wurden, während der Grundfrequenzverlauf zwischen den jeweiligen Partnern in den Paaren gemorpht wurde, und zwar am 50%-Punkt, also in der Mitte zwischen den Grundfrequenzverläufen der Originalsignale. So entstanden also Endpunkte, die jeweils den gleichen Grundfrequenzverlauf hatten; diese Eigenschaft galt also folglich auch für die Zwischenschritte in den nun zu erzeugenden Kontinua.

Für diese wurden alle Eigenschaften zwischen den künstlichen Endpunkten mit 10 Schritten gemorpht (wobei hier eigentlich nur die spektralen Eigenschaften interessieren, da die Ausgangssignale vor der Analyse in *praat* auf jeweils 70 dB skaliert worden waren und temporale Eigenschaften bei der Identifikation von Paarpartnern, die phonologische Länge jeweils teilen, keine allzu große Rolle spielen sollte), so dass für jedes der beiden Kontinua elf Stimuli entstanden. Um festzustellen, welche spektralen Eigenschaften in den Zwischenschritten vorherrschen, wurden beide Kontinua wiederum akustisch mit *forest* (bzw. *f0ana* für die Grundfrequenz) ausgemessen. Wie die Abbildungen 4.3 und 4.4 zeigen, sind die durch die Morphingmethode entstandenen Stimuli hauptsächlich dadurch gekennzeichnet, dass F1 von einem niedrigeren Wert zu einem höheren Wert wandert, und zwar in

linearem Verlauf (die Abweichungen vom linearen Verlauf sind, ähnlich wie die unsystematische Variation in F2 und F3, Messfehlern geschuldet; siehe auch die leichte Variation in den *f0*-Messwerten). Der jeweils rechte Teil der Abbildungen zeigt zudem, dass erfolgreich der *F1-f0*-Abstand, gemessen in Bark, systematisch und ebenfalls praktisch linear variiert wurde.

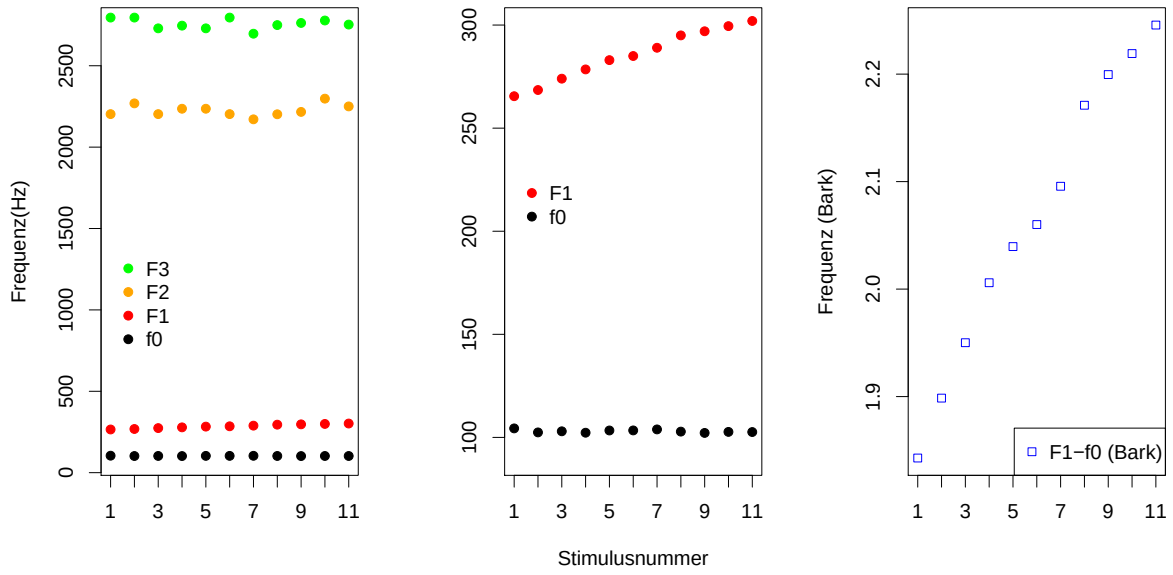

Abbildung 4.3: *f0 und Formanten (in Hertz) im ersten Vokal der Stimuli aus dem* bieten-beten*-Kontinuum des Vortests. Die Grundfrequenz wurde unverändert resynthetisiert, und ihr Verlauf entspricht dem 50%-Morphing-Punkt zwischen der f0 von bieten und der f0 von beten. Die Formantänderungen entstehen durch Morphing der Spektra von* /iː/ *nach* eː *in* bieten *und* beten*. Die Abbildung links zeigt die gemessenen Grundfrequenzen und Formantwerte (F1-3) der so erzeugten Stimuli, die mittlere Abbildung ausschließlich Grundfrequenz und F1; die rechte Abbildung zeigt den F1-f0-Abstand in Bark.*

4.2.6 Perzeptueller Vortest

Die auf die beschriebene Weise entstandenen Kontinua wurden in einen vorher ebenfalls auf im Mittel 70 dB skalierten Trägersatz, *Ich habe ... gesagt*, der einer *Ich habe bäten gesagt*-Äußerung entnommen worden war, per Konkatenation in *praat* eingefügt. Die so entstandenen Sätze wurden mittels eines ExperimentMFC-Scripts in *praat* 8 Versuchspersonen vorgeführt.

Die 8 Versuchspersonen, 5 Frauen und 3 Männer im Alter zwischen 22 und 35 Jahren, waren Sprecher des Standarddeutschen, wobei zwei (ein Mann und eine Frau) leichte Ein-

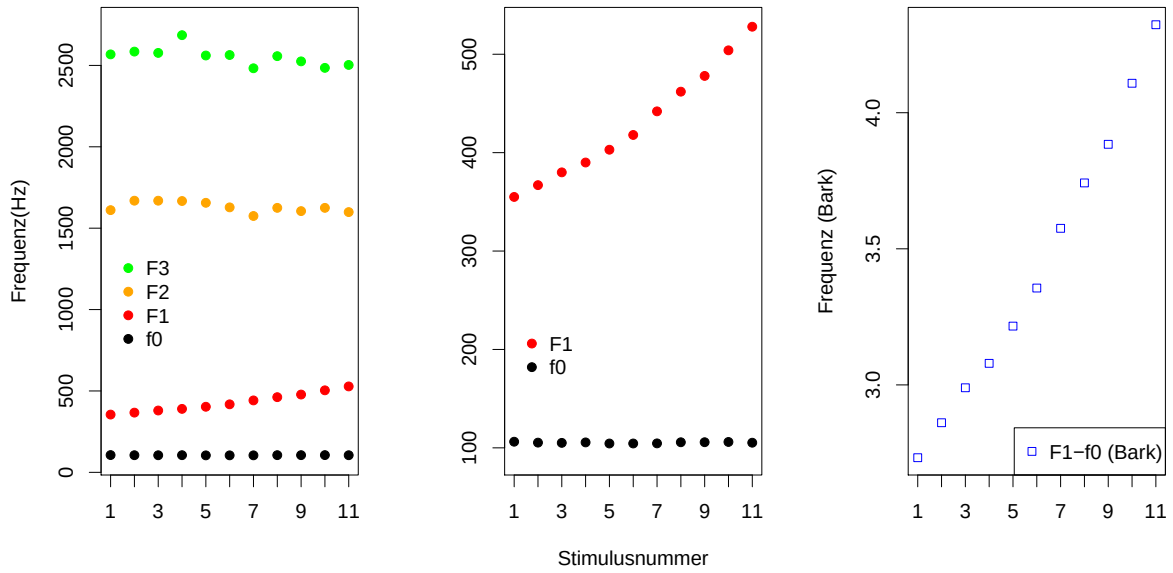

Abbildung 4.4: *f0 und Formanten (in Hertz) im ersten Vokal der Stimuli aus dem* bitten-betten*-Kontinuum des Vortests. Die Abbildung links zeigt die gemessenen Grundfrequenzen und Formantwerte (F1-3) der durch Morphing erzeugten Stimuli, die mittlere Abbildung ausschließlich Grundfrequenz und F1; die rechte Abbildung zeigt den F1-f0-Abstand in Bark.*

flüsse der münchener Variante der bairisch gefärbten Umgangssprache zeigten, was für tolerierbar gehalten wurde, da die Versuchspersonen im nachfolgenden Experiment ebenfalls schwer bezüglich der Standardlautung ihres Sprachgebrauchs überprüft werden konnten. Alle acht Versuchspersonen für dieses Vorexperiment wurden am Institut für Phonetik und Sprachverarbeitung rekrutiert.

Das Experiment war ein forced-choice-Identifikationsexperiment mit folgendem Ablauf: Die Versuchspersonen hörten in randomisierter Reihenfolge einen der Sätze aus einem der beiden Kontinua und mussten, abhängig vom Kontinuum, einen der beiden angebotenen Buttons anklicken, wovon sich einer auf der linken, einer auf der rechten Bildschirmseite befand. Für das *Ich habe bieten gesagt - Ich habe beten gesagt*-Kontinuum waren die Antwortmöglichkeiten *bieten* bzw. *beten*, für das ungespannt-Kontinuum entsprechend *bitten* bzw. *betten*; die Reihenfolge der Antwortmöglichkeiten war ausbalanciert, es befand sich also z. B. *beten* in 50% der Fälle links, in der anderen Hälfte der Fälle rechts auf dem Bildschirm. Über den Antwortmöglichkeiten stand die Frage *Was hat er gesagt?*.

Der Sinn dieses Vortest war, wie erwähnt, in beiden Kontinua jeweils jenen Stimulus ausfindig zu machen, der ambig war, um die Chancen für eine Beeinflussbarkeit durch Grundfrequenzvariation zu erhöhen. Hierzu wurde jener Stimulus ausgemacht, dessen gemittelte proportionale Kategorisierung am nächsten bei 0,5 lag. Hierzu wurden, wie bei Kategorisierungsexperimenten üblich, für jedes Kontinuum eine S-förmige Identifikations-

kurve sowie der darin befindliche kategoriale Umkipppunkt ermittelt, um dem Leser eine Idee davon zu geben, wie die Kontinua aufgeteilt wurden.

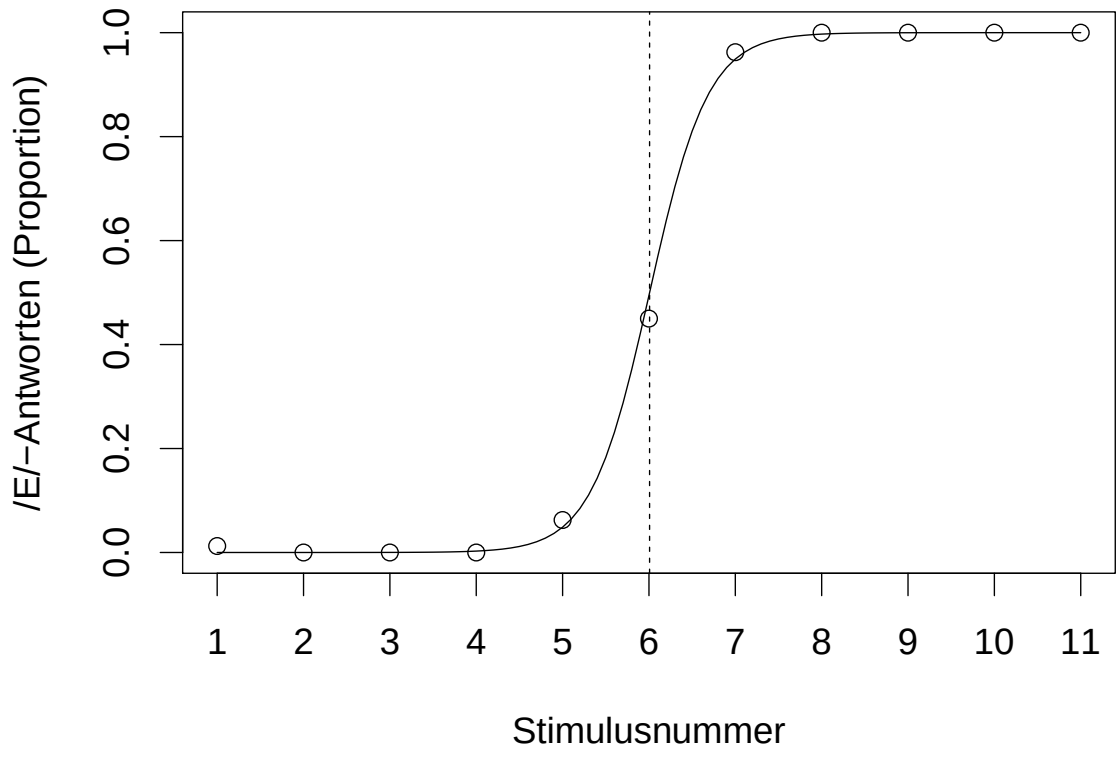

Abbildung 4.5: *Rohdaten (über die acht teilnehmenden Versuchspersonen gemittelte Proportionen), überlagerte Identifikationskurve und gemittelter Umkipppunkt für das* bitten-betten-*Kontinuum, eingebettet in die Äußerung* Ich habe ... gesagt.

Wie ein Vergleich der Abbildungen 4.5 und 4.6 zeigt, unterscheiden sich die Aufteilungen für die beiden Kontinua, auch wenn in beiden Fällen natürlich in einem Generalized Linear Mixed Model der Faktor Stimulusnummer die Antworten signifikant beeinflusst (*bieten-beten*-Kontinuum: $\chi^2[1] = 74,0; p < 0,001$, *bitten-betten*-Kontinuum: $\chi^2[1] = 59,7; p < 0,001$). Während für *bitten-betten* die Stimuli 1-5 zu mehr als 90% als *bitten* und die Stimuli 7-11 ebenfalls zu mehr als 90% als *betten* wahrgenommen werden, und somit nur Stimulus 6 ambig ist (proportionale Bewertung: 0.45), wird bei *bieten-beten* der *beten*-Endpunkt nur zu 63,75% als *beten* kategorisiert. Der Stimulus mit einer proportionalen Einschätzung nächst zu 0.5 ist Stimulus 10 (0.4625).

Diese beiden Stimuli waren die Grundlage für das Hauptexperiments diesen Kapitels,

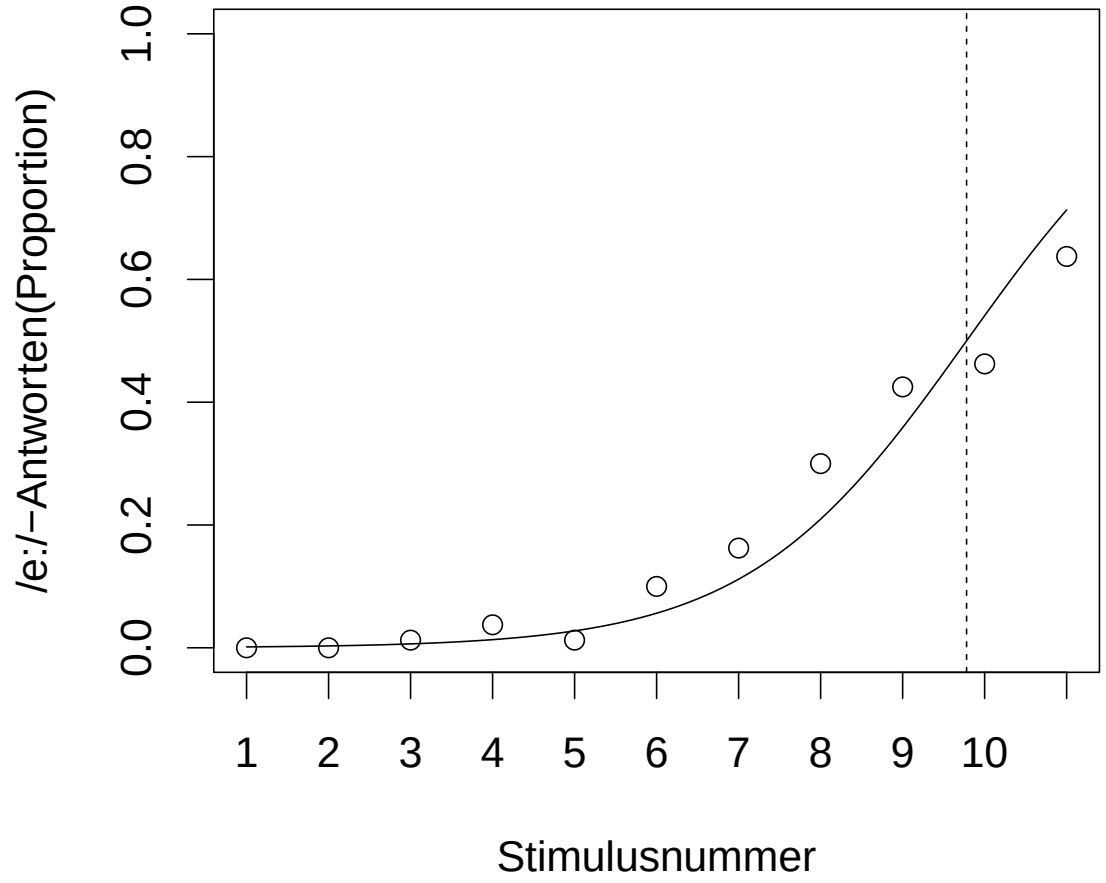

Abbildung 4.6: *Rohdaten (über die acht teilnehmenden Versuchspersonen gemittelte Proportionen), überlagerte Identifikationskurve und gemittelter Umkipppunkt für das* bieten-beten-*Kontinuum, eingebettet in die Äußerung* Ich habe ... gesagt.

das lediglich durch *f0*-Variation die Kategorisierung in zwei vokalhöhenverschiedene Vokale - /iː/ und /eː/ sowie /ɪ/ und /ɛ/ - zu beeinflussen versucht.[1]

[1]Da dieser Vortest nur dazu diente, spektral ambige Stimuli zu finden, um diese dann für ein weiteres Experiment, in dem die Grundfrequenz manipuliert werden soll, zu benutzen, ist es wenig zielführend, ausführlich mögliche Gründe für die Unterschiede in der Kategorisierung beider Stimulikontinua zu diskutieren. Dennoch hier eine Anmerkung, die Prof. Dr. Harrington in persönlicher Kommunikation als Vermutung geäußert hat: Der Sprecher war, wie erwähnt, Sprecher der norddeutschen Variante des Standarddeutschen. Die Sprecher dieser Variante neigen möglicherweise dazu, halbgeschlossene Vokale in Richtung der

Stimulus	f0 (Hz)	F1 (Hz)	F2 (Hz)	F3 (Hz)
bitten-betten-6	105	418	1628	2564
bieten-beten-10	103	300	2298	2778

Tabelle 4.1: *Akustische Eigenschaften (Grundfrequenz und die ersten drei Formantwerte) ambiger Stimuli, die als Ausgangsmaterial für das nächste Experiment benutzt wurden. Die Werte wurden als Medianwerte der mittleren 20% der Vokaldauer ermittelt.*

4.2.7 Erzeugung je zweier *bieten-beten*- und *bitten-betten*-Kontinua durch Grundfrequenzmanipulation

Für beide Wortpaare - *bieten-beten* und *bitten-betten* - wurden je zwei Kontinua, in denen die Grundfrequenzvariation zur Vokalklassifikation beitragen sollte, erstellt. In je einem Kontinuum wurde die Grundfrequenz nur in der Silbe, deren Nucleus der zu beeinflussende Vokal war, manipuliert, in je einem zweiten wurde die Grundfrequenz global manipuliert. Im jeweils ersten Kontinuum führt die *lokale* Variation zusätzlich auch zu einer gewissen Variation der Art der Satzakzentuierung, im jeweils zweiten Kontinuum führen die *globalen* Änderung zu einer Variation des Registers des Sprechers.

Die beiden genannten ambigen Stimuli aus dem Vorexperiment mit den in Tabelle 4.1 genannten akustischen Eigenschaften wurden in *praat* manipuliert. Hierzu wurden die Klangdateien eingelesen und mittels der *to manipulation*-Methode bearbeitet. Dabei kommt die Pitch Synchronous OverLap-Add-(PSOLA)-Methode, wie sie in Moulines und Charpentier (1990) beschrieben ist, zum Einsatz. Die Implementierung in praat erlaubt es, Signalabschnitte auszuwählen und in diesen - neben der Dauer - die Grundfrequenz in selbst zu wählenden Einheiten zu manipulieren. Im vorliegenden Fall wurde als Einheit der Halbton gewählt. Ziel der Manipulation war es, relativ eindeutige *beten*- und *bieten*-Exemplare zu erzeugen. Da die Formanten davon unabhängig ja weiterhin als auditiver Cue ambig sein würden, musste die Manipulation der Grundfrequenz einen relativ weiten Bereich umspannen, ohne zu unnatürlich zu werden. Nach einigen Versuchen kam der Autor zu dem Schluss, dass für die vorliegende Sprecherstimme hierzu ein Abstand (zwischen den /iː/- und /eː/-Enden des Kontinuum) von 5 Halbtönen nötig sei. Da pro Kontinuum elf Stufen angestrebt wurden, ergab sich so eine Schrittweite von einem halben Halbtonschritt von einem Stimulus zum nächsten. Tabelle 4.3 zeigt die gerundeten durchschnittlichen Grundfrequenzwerte in Hz in der /bʊ/-Silbe. Wie zu sehen ist, entspricht eine Änderung um einen halben Halbton in etwa einer Änderung um 3 Hz, mit etwas größeren Werten in höherer Lage, und etwas tieferen Werten in tiefer Lage. Laut Zwicker (1982, Seite 55) entspricht dies in etwa den Werten für die „eben wahrnehmbare Frequenzänderung“ also

geschlossenen zu verschieben, zumindest jene der gespannten Varianten. Dies könnte erklären, warum die zumeist süddeutschen Hörer dazu neigen, vermehrt /iː/ wahrzunehmen, während das /ɪ/-/ɛ/-Kontinuum fast perfekt kategorial wahrgenommen wird.

der *just noticeable differenz* (jnd) für Frequenzunterschiede bei Sinustönen, die er mit 3.6 Hz für den Frequenzbereich von 0 bis circa 500 Hz angibt; laut Klatt (1973) ist die jnd für Grundfrequenzunterschiede in den komplexeren Sprachsignalen allerdings geringer und wird mit circa 2 Hz angegeben.

Wie der Literatur zu entnehmen ist, und auch an der Verteilung der intrinsischen Grundfrequenz des vorliegenden Sprechers in Kapitel 4.2.2 und dort in Abbildung 4.2 zu sehen ist, gehen fünf Halbtöne weit über das hinaus, was für das Deutsche im Allgemeinen und für den hier benutzten Sprecher im besonderen als Unterschied zwischen in der Vokalhöhendimension benachbarten Vokalkategorien zu erwarten ist, sondern entspricht vielmehr ziemlich exakt dem Gesamtbereich vom maximalen *If0*-Wert bei einem /ɪ/ - ca. 119 Hz - bis zum Minimum um 88 Hz bei einem /aː/ des gegebenen Sprechers, was 5.2 Halbtönen entspricht.

Wie in Kapitel 4.2.6 beschrieben, entstanden die Grundfrequenzverläufe der Kontinua des Vortests durch Morphing der Verläufe der /iː/-/eː/- bzw. /ɪ/-/ɛ/-Paare. Daher wurden die neuen Kontinua so erstellt, dass die im Vortest als ambig bewerteten Stimuli der jeweilige Mittelpunkt der neuen Kontinua sein sollten. Von diesen ausgehend, wurde die Grundfrequenz in einer Schrittweite von je einem halben Halbtonschritt (bzw. um 50 cents) nach oben und unten verschoben, bis als Endpunkte zweieinhalb Halbtöne vom Ausgangspunkt entfernt liegende Tokens entstanden. Die so entstandenen, fünf Halbtöne auseinanderliegenden Tokens wurden vom Autor und einer weiteren Phonetikerin begutachtet und als /iː/,/eː/,/ɪ/ und /ɛ/ kategorisiert.

Pro Wortpaar (*bieten-beten* und *bitten-betten*) wurden zwei Manipulationsarten verwendet. In der ersten Variante wurde ausschließlich die Grundfrequenz in der akzentuierten Silbe verändert, also in der ersten Silbe der ursprünglich ambigen *bieten/beten* bzw. *bitten/betten*-Wörter, die in den Trägersatz *Ich habe ZIELWORT gesagt* eingebettet waren. Hier wurde also nicht die globale Lage - man könnte auch Register sagen - der Stimme variiert, sondern nur eine Silbe der Äußerung, wobei die *lokale* Variation natürlich auch eine gewisse Variation der Art der Akzentuierung zur Folge hatte. Diese wurde durch eine trainierte Phonetikerin nach GTobi gelabelt; diese Etikettierung ist in Tabelle 4.2 zu finden.

Kontinuum	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
bitten-betten	LH*	LH*	LH*	LH*	H*	H*	H*	H*	L*H	L*H	L*H
bieten-beten	LH*	LH*	LH*	LH*	H*	H*	H*	H*	L*	L*	L*

Tabelle 4.2: *Etikettierung der lokal grundfrequenzmanipulierten (aber immer akzentuierten) Silbe /bV/ in* Ich habe b/V/ten gesagt-*Stimuli, mit bVten=bitten-betten oder bVten=bieten-beten. Der nachfolgende Verlauf* ten gesagt *wurde stets mit* L-% *gelabelt. Die Form in den* globalen *Kontinua entspricht stets der Form* H*, *da diese Grundfrequenzverläufe bis auf die globale Lageveränderung identisch sind mit dem der jeweils sechsten Stimuli im* lokal-*Kontinuum.*

In der zweiten Variante wurde der gesamte Grundfrequenzverlauf der Äußerung *global* verschoben. Durch Messungen in den akzenttragenden Silben konnte bestägigt werden, dass beide Varianten - globale und lokale Grundfrequenzverschiebung - die gleichen Auswirkungen in diesen Silben hatte. Die Werte in diesen Silben, zurückgerechnet in Hertz, sind in Tabelle 4.3 zu finden.

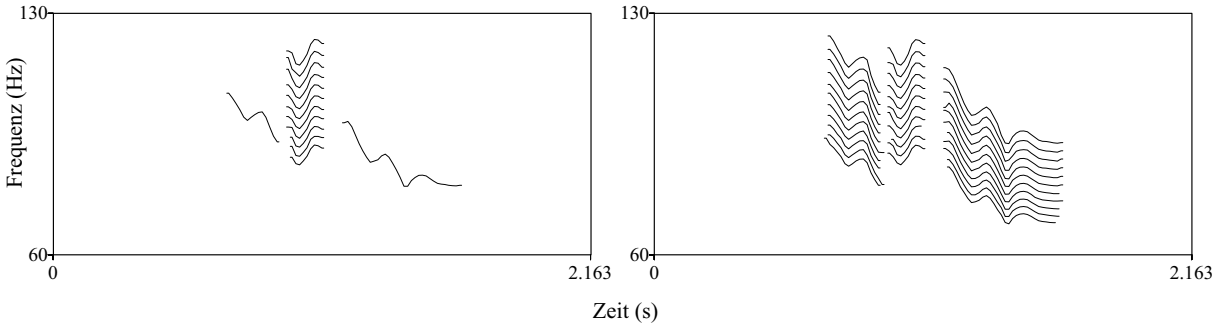

Abbildung 4.7: *f0-Verlauf von bieten-beten in den zwei Typen der Verschiebung, lokal (links) und global (rechts), als Beispiel für lokale und globale Grundfrequenzverschiebung.*

Kontinuum	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
bitten-betten	121	117	114	111	108	105	102	99	96	93	91
bieten-beten	119	116	112	109	106	103	100	97	94	91	89

Tabelle 4.3: *Grundfrequenzkontinua für die Resynthesen von* bitten-betten *und* bieten-beten*; Werte in Hertz, gemessen mit* praat*. Stimulus 6 entspricht in beiden Kontinua dem ambigen Stimulus aus dem Vortest 4.2.6; Formantwerte siehe in Tabelle 4.1. Von diesem Stimulus ausgehend, wurde* f0 *mit einer Schrittweite von einem halben Halbton nach oben (Richtung Stimulus 1) und unten (Richtung Stimulus 11) manipuliert. In beiden Kontinua variiert die Grundfrequenz somit um 5 Halbtöne. Berechnet man den Abstand in Bark zwischen den jeweiligen* F1*-Werten und den hier präsentierten Grundfrequenzwerten, ergibt sich eine Variation zwischen 2.0 Bark und 2.37 Bark für das* bitten-betten*-Kontinuum, und von 3.18 Bark bis 3.55 Bark für das* bieten-beten*-Kontinuum.*

Wie zu sehen ist, sind die Werte in beiden Kontinua, da ihre jeweiligen Ausgangswerte (jeweils bei Stimulus 6) mit 105 Hz und 103 Hz recht nahe beieinander lagen, sehr ähnlich, wenn der Unterschied, in Halbtönen berechnet, doch immerhin circa einen viertel Halbton und damit eine halbe Schrittweite beträgt. Da beide Kontinua wegen der durch die Gespanntheitsopposition gegebenen unterschiedlichen Vokaldauern (*bieten/beten*: 129 ms, *bitten/betten*: 70 ms) ohnehin nur eingeschränkt vergleichbar sind, wurde darauf verzichtet, sie vollständig aneinander anzugleichen. Desweiteren fällt auf, dass die Wertebereiche in

etwa demjenigen der intrinsischen Grundfrequenz für die gesamte Vokalhöhenvariationsbreite (bei diesem Sprecher von /aː/ zu /ɪ/) entsprechen. Diese Beinahe-Übereinstimmung war nicht intendiert, sondern ein Zufallsprodukt, entstanden durch die Auswahl der ursrünglichen Tokens aus den Aufnahmen des Sprechers für die Kontinua des Vortests.

4.2.8 Präsentation in dem webbasierten Perzeptionsexperimenttool Percy

Die auf die beschriebene Weise entstandenen vier Kontinua - *bieten-beten* lokal, *bieten-beten* global, *bitten-betten* lokal und *bitten-betten* global, wurden mittels des webbasierten Experimenttools Percy, entwickelt von PD Dr. Christoph Draxler (Draxler, 2011), in das Internet gestellt; Versuchspersonen wurden mittels eines durch den Infodienst der Ludwig-Maximilians-Universität (LMU) versendeten elektronischen Rundschreibens unter der Studentenschaft der LMU gesucht. Den Teilnehmern wurde bei Abschluss der Teilnahme am Experiment die Teilnahme an einem Preisausschreiben mit der Gewinnmöglichkeit über 50 Euro versprochen. Auf diese Weise konnten 126 Versuchspersonen rekrutiert werden.

Diesem Vorteil einer großen Anzahl von Versuchspersonen steht jedoch der Nachteil gegenüber, dass eine Kontrolle der Auswahl der Versuchspersonen nur eingeschränkt und im Nachhinein möglich ist, d.h. über deren Selbstauskunft, die während der Anmeldung zum Experiment abgefragt wird. Den Versuchspersonen steht z. B. frei, ob sie Kopfhörer oder Lautsprecher benutzen. Unabhängig davon, dass keine Kontrolle darüber besteht, welche Art von Klangwiedergabegeräten benutzt wurde, kann doch zumindest untersucht werden, ob grundsätzlich zwischen den Antworten der Versuchspersonen, die Lautsprecher verwendeten, und derer, die Kopfhörer benutzten, Unterschiede bestehen.

Den Versuchspersonen wurde in Form eines two-alternative forced-choice Perzeptionsexperiments abgefragt, ob sie einen gegebenen Stimulus, der nur einmal wiederholt werden konnte, entweder als *bieten* oder *beten*, bzw. als *bitten* oder *betten* wahrgenommen hatten. Die beiden Kontinua wurden getrennt voneinander präsentiert (*bieten-beten* zuerst); innerhalb der Kontinua wurden die Stimuli randomisiert und mit sechs Wiederholungen pro Stimulus dargeboten. Die Antwortmöglichkeiten waren auch in diesem Experiment ausbalanciert präsentiert. In dem einleitenden Test zum Experiment wurden die Hörer ausdrücklich darauf aufmerksam gemacht, dass es sich um die Äußerungen *eines* Sprechers handelt: „Wenn Sie auf den folgenden Seiten auf das Lautsprechersymbol klicken, dann werden Sie einen Sprecher hören, der immer den Satz *Ich habe ZIELWORT gesagt* wiederholt, wobei das ZIELWORT variiert. Ihre Aufgabe ist es, anzugeben, welches Wort der Sprecher Ihrer Meinung nach gesagt hat“.

Generalisierte Lineare Gemischte Modelle wurde angewendet, um die binominalen Antworten der Sprecher auszuwerten. Zunächst wurde ein Modell über alle vier Kontinua unter Ausklammerung der Sprechervariation und mit den fixed factors *Stimulus* und *Darbietung* (dieser Zwischen-Subjekt-Faktor mit den zwei Stufen *Kopfhörer* und *Lautsprecher*) berechnet, um festzustellen, ob die Daten derjenigen Sprecher, die Lautsprecher verwendet hatten, überhaupt in die Analyse einfließen sollten oder nicht. *Darbietung* hatte aber

einen signifikanten Einfluss auf die Antworten ($\chi^2[1] = 5,88; p < 0,05$, siehe auch Abbildung B.1 im Anhang auf Seite 210), so dass beschlossen werden musste, die Sprecher in zwei Gruppen, nämlich in die 59 Kopfhörerbenutzer und in die 67 Lautsprecherbenutzer, aufzuteilen. Wie sich herausstellte, waren die Antworten der Lautsprecherbenutzer - im Gegensatz zu denen der Kopfhörerbenutzer - in allen vier Kontinua nur bedingt beeinflusst von der Grundfrequenzvariation der Stimuli, so dass beschlossen wurde, im Hauptteil dieser Arbeit nur die Antworten der Kopfhörerbenutzer auszuwerten und die Analyse der Lautsprecherbenutzer im Anhang zu präsentieren. Diese Auswahl erfolgte auch, da der Vortest, in dem die ambigen Stimuli ermittelt worden waren, die Grundlage für das vorliegende Experiment bildeten, ebenfalls mit Kopfhörern durchgeführt worden war, d.h. der bias bei den Lautsprecherbenutzern, der in allen vier Fällen in Richtung der im Vokalraum sich höher befindenden Variante (also /iː/ bzw. /ɪ/) ging, mag auch dadurch gesetzt worden sein, dass die Stimuli - über Lautsprecher präsentiert - möglicherweise nicht spektral ambig erschienen, sondern diesen bias bereits enthielten.

Die 59 Kopfhörerbenutzer, deren Antworten im folgenden ausgewertet werden, waren zwischen 17 und 40 Jahren alt (Durchschnittsalter: 23.3 Jahre). Die 45 Frauen und 14 Männer kamen aus folgenden Bundesländern bzw. Regionen: 44 aus Bayern, 5 aus Baden-Württemberg, je zwei aus Nordrhein-Westfalen und Sachsen, und je ein Teilnehmer aus Hessen, Niedersachsen, Schleswig-Holstein, dem Saarland, Luxemburg, sowie Brandenburg. Die Versuchspersonen wurden gebeten, eine Selbstauskunft über den von Ihnen gesprochenen Dialekt zu geben. So gaben 31 Hörer Standarddeutsch an, 13 bairische und 5 ostmittelfränkische Dialektformen, 8 gaben an, eine Form aus dem Bereich alemannischer Dialekte zu sprechen; außerdem bewertete sich eine Versuchsperson als Sächsisch-Sprecher und eine als Sprecher der im Westfälischen gesprochenen Dialekte. Da bezüglich möglicher dialektal beeinflusster Unterschiede in der Nutzung des Cues intrinsischer Grundfrequenz keine Hypothesen aufgestellt worden waren, hielt der Autor diese impressionistische Selbstauskunft für hinreichend.

4.3 Ergebnisse

Für alle vier Kontinua, also bieten-beten-global, bieten-beten-lokal, bitten-betten-global und bitten-betten-lokal wurde Generalisierte Lineare Gemischte Modelle mit den propotionalen Antworten der Hörer als abhängiger Variable, dem Kontinuum als fixed factor und unter Ausklammerung der Hörer als unerwünschten Variationsbringern berechnet, die ergaben, dass alle vier Wertevariationen durch die Stimulusnummer erklärbar waren (bieten-beten-lokal: $\chi^2[1] = 55,4; p < 0,001$, bieten-beten-global: $\chi^2[1] = 49,6; p < 0,001$, bitten-betten-lokal: $\chi^2[1] = 57,5; p < 0,001$, bitten-betten-global: $\chi^2[1] = 48,6; p < 0,001$). Desweiteren ergab sich bei zusätzlicher Modellierung aller Daten einer Gespanntheitsstufe, die zusammengelegt wurden, und unter Einführung eines weiteren Faktors, nämlich *Typ der Verschiebung* (mit den Stufen *global* und *lokal*), weder zwischen den zwei *gespannten* Kontinua ($\chi^2[1] = 3,8; n.s.$) noch zwischen den zwei *ungespannten* Kontinua ($\chi^2[1] = 3,4; n.s.$) ein signifikanter Unterschied (siehe hierzu Abbildung 4.10 a)), d.h. es macht keinen Unter-

schied, ob die Grundfrequenz *lokal* oder *global* verschoben wurde.

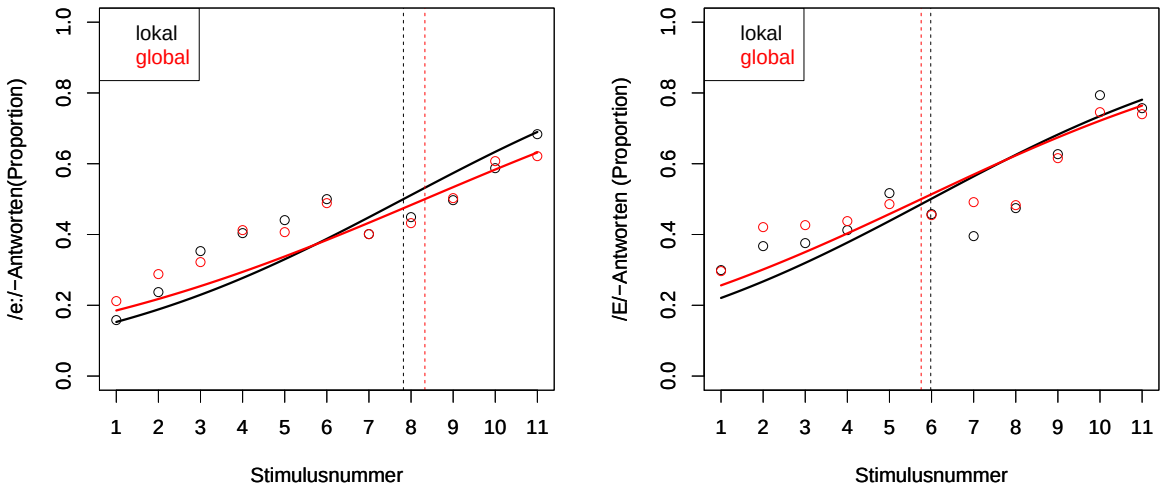

Abbildung 4.8: *Identifikationskurven zu je zwei Kontinua von* Ich habe *bieten* gesagt *zu* Ich habe *beten* gesagt *(linke Abbildung) bzw. von* Ich habe *bitten* gesagt *zu* Ich habe *betten* gesagt*, wobei die Grundfrequenz von Stimulus 1 bis Stimulus 11 um fünf Halbtöne* lokal *(schwarz) bzw.* global *(rot) variiert wurde.*

Man kann aus diesen Daten die über alle Versuchspersonen gemittelten proportionalen Identifikationswerte pro Stimulus errechnen und abbilden, so wie in Abbildung 4.8; überlagert sind in dieser Abbildung die modellierten Identifikationskurven sowie die mittleren Umkipppunkte, letztere in Gestalt gestrichelter vertikaler Linien. Tatsächlich werden aber die Daten eines jeden einzelnen Sprechers modelliert, wodurch sich für jeden Sprecher ein Umkipppunkt und ein Wert für die Steigung der Kurve ergibt. Die Steigungswerte aller Sprecher sind in den oberen beiden Boxen der Abbildung 4.10 b) abgebildet. Wie zu sehen, gibt es offenbar Hörer, deren Identifikationskurve sehr flach wird oder sogar negative Steigungswerte annimmt. Daraus ergibt sich für die Umkipppunkte, - obschon diese über alle Hörer gemittelt im Bereich zwischen Stimulus 1 und 11 liegen - dass einige Hörer gar keinen Umkipppunkt im Bereich des Kontinuums haben, sondern stattdessen nur theoretische jenseits der Grenzen des Kontinuum, d.h. sie können nicht den Cue der Intrinsischen Grundfrequenz nutzen und hören deshalb keine systematische Unterschiede zwischen den einzelnen Stimuli. Diese theoretischen „Umkipppunkte“ können Werte annehmen, die sehr weit von den Werten im Kontinuum entfernt liegen, so dass sie die Normalverteiltheit der Umkipppunktwerte stören. Abgesehen davon, sind diese Werte außerhalb des Kontinuumbereichs natürlich nicht interpretierbar. Daher wurden diese Werte aus der weiteren Analyse entfernt.

Hierzu wurde folgendermaßen vorgegangen: Da gepaarte t-Tests benutzt wurden, um Unterschiede zwischen den beiden Kontinua pro *Gespanntheit*sstufe zu ermitteln (also ein Wert für das *global*-, ein Wert für das *lokal*-Kontinuum), mussten gegebenenfalls beide Werte entfernt werden, auch wenn nur ein Umkipppunkt kleiner als 1 oder größer als 11 war. Bei bieten-beten in

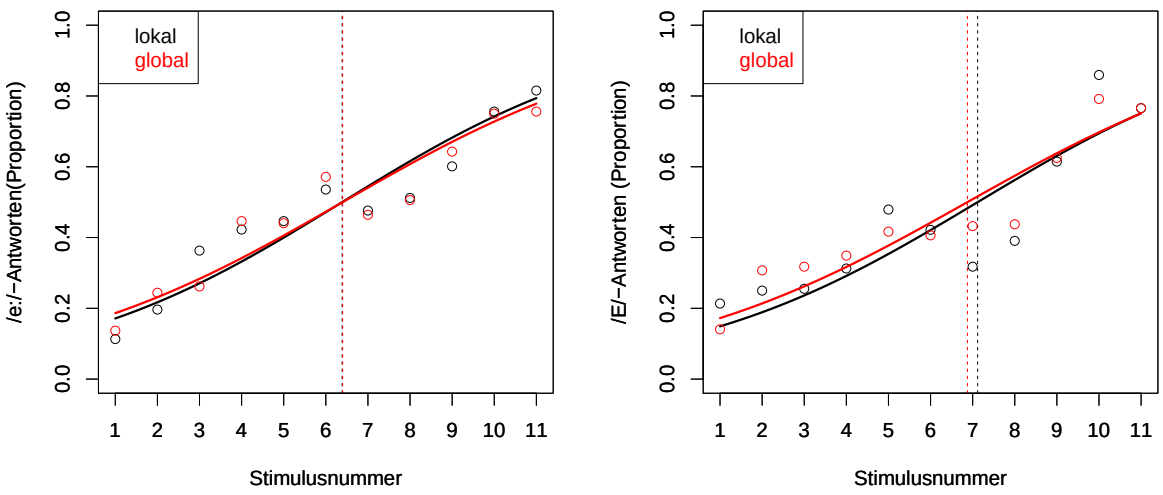

Abbildung 4.9: *Identifikationskurven zu je zwei Kontinua von* Ich habe *bieten* gesagt *zu* Ich habe *beten* gesagt *(linke Abbildung) bzw. von* Ich habe *bitten* gesagt *zu* Ich habe *betten* gesagt*, wobei die Grundfrequenz von Stimulus 1 bis Stimulus 11 um fünf Halbtöne* lokal *(schwarz) bzw.* global *(rot) variiert wurde. Hier sind, im Unterschied zu Abbildung 4.8, nur die Daten der Versuchspersonen abgebildet, die einen Umkipppunkt aufwiesen (29 im gespannt-Kontinuum, 32 im ungespannt-Kontinuum).*

beiden *Typen der Verschiebung* betraf dies die Wertepaare von 30 Versuchspersonen, bei denen in mindestens einem der beiden Kontinua Werte jenseits der Kontinuumswerte auftraten. Bei bitten-betten betraf dies die Umkipppunkte von 27 Versuchspersonen. Diese 30 bzw. 27 Versuchspersonen konnten also in mindestens einem der zwei Kontinua pro Gespanntheitsstufe *f0* nicht hinreichend nutzen, um zwischen /iː/ und /eː/ bzw. zwischen /ɪ/ und /ɛ/ zu unterscheiden. In beiden Fällen ergaben die erwähnten gepaarten t-Tests für jene Sprecher mit Umkipppunkt im Kontinuumsbereich keine signifikanten Unterschiede (bieten-beten: $t[27] = 0,03; n.s.$; bitten-betten: $t[31] = 0,33; n.s.$, siehe Abbildung 4.10 b), untere zwei Boxen; die in der Abbildung 4.8 dargestellten mittleren Umkipppunkte beruhen aber auf den Daten aller 59 Sprecher; ein Vergleich der Abbildung aller Versuchspersonen mit denen, die einen Umkipppunkt aufweisen, ist gegeben durch die Abbildungen 4.8 und 4.9).

Was die Steigungswerte angeht, wurden die Daten aller 59 Personen benutzt, um gepaarte t-Tests zu berechnen: in beiden Gespanntheitsstufen ergaben sich signifikante Unterschiede zwischen *globaler* und *lokaler* Grundfrequenzverschiebung, bei bieten-beten $t[58] = 3,38; p < 0,005$, und bei bitten-betten $t[58] = 2,04; p < 0,05$. Die Unsicherheit der Kategorisierung ist also jeweils in der *globalen* Variante höher, da dort die niedrigeren Steigungswerte auftreten (siehe Abbildung 4.10 b), obere Boxen). Berechnet man stattdessen eine Varianzanalyse mit Messwiederholung über alle vier Kontinua mit der abhängigen Variable *Steigung*, den zwei Inner-Subjekt-Faktoren *Typ der Verschiebung* und *Gespanntheit* und den Hörern als auszuklammernden Faktor, so ergibt sich eine signifikanter Effekt nur für *Typ der Verschiebung* ($F[1, 58] = 14,6; p < 0,05$), aber

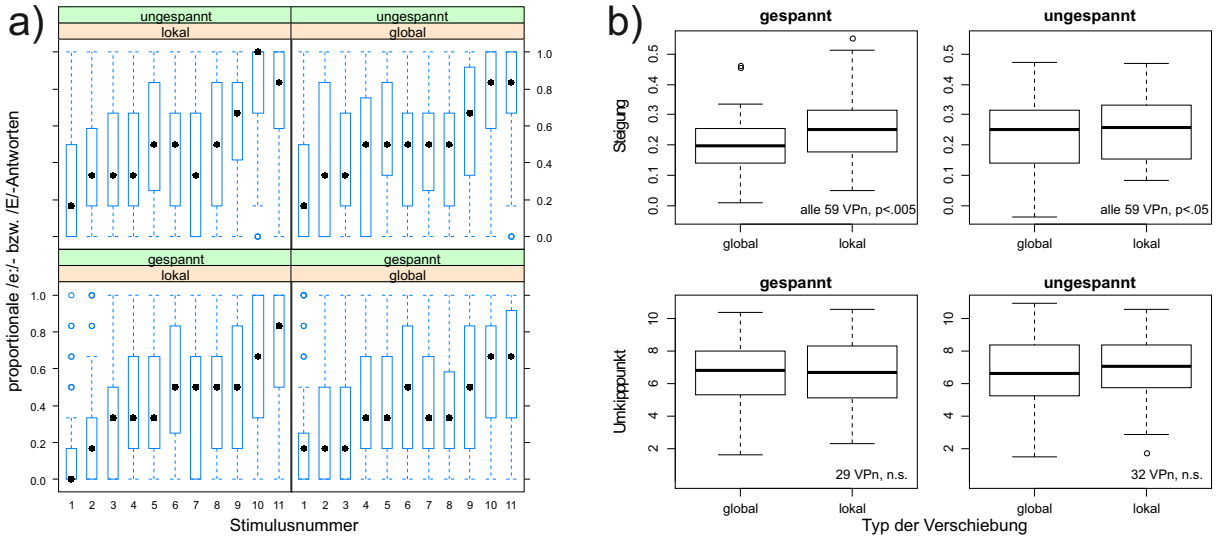

Abbildung 4.10: *a): Kontinuaaufteilung unter dem Einfluss der Gespanntheit (oben: ungespannt; unten: gespannt) und des Verschiebungstyps (links: lokal; rechts: global). b): Verteilungen der Steigungen der Identifikationskurven aus Abbildung 4.8 für alle 59 Versuchspersonen, sowie die Verteilungen der Umkipppunkte jener 29 (Kontinua mit dem* gespannt*-Vokalpaar) bzw. 32 (Kontinua mit dem* ungespannt*-Vokalpaar) Versuchspersonen, die in jeweils beiden Kontinua einen Umkipppunkt im eigentlichen Sinne, also im Bereich zwischen Stimulus 1 und Stimulus 11, hatten. Die* p*-Wertangaben beruhen auf den Ergebnissen gepaarter* t*-Tests, siehe Text.*

weder für *Gespanntheit* ($F[1, 58] = 0, 6; n.s.$) noch für die Interaktion beider Faktoren ($F[1, 58] = 0, 98; n.s.$). Post-hoc durchgeführte Bonferroni-korrigierte gepaarte t-Tests ergeben dann die für die pro *Gespanntheit* durchgeführten Vergleiche des Einflusses von *Typ der Verschiebung* auf die Steigungen jeweils gleichen t-Werte wie oben berichtet, aber wegen der Bonferroni-Korrektur der p-Werte sind nur noch in den *gespannten* Kontinua *global-lokal*-Unterschiede als signifikant zu betrachten ($t[58] = 3, 38; p < 0.01$; im *bitten-betten*-Fall: $t[58] = 2, 04; n.s.$). Die *Gespanntheit* hat in keinem *Typ*-Paar einen Einfluss (global: $t[58] = 1, 1; n.s.$; lokal: $t[58] = 0, 1; n.s.$)

Alle Vergleiche über die *Gespanntheit* hinweg sind etwas fragwürdig, da die Kontinua zwar sehr ähnlich, aber nicht identisch sind, wie bereits in 4.3 erwähnt; beide Kontinua umfassen 5 Halbtöne vom ersten bis zum elften Stimulus bei einer Schrittweite von einen halben Halbton, aber das *bitten-betten*-Kontinuum liegt circa einen viertel Halbton - und damit eine halbe Schrittweite - höher. Dennoch wollen wir den Einfluss der *Gespanntheit* untersuchen, wobei wir nicht nur die Hörervariation, sondern auch die durch den Faktor *Typ der Verschiebung* bedingte Variation ausklammern wollen. Hierzu verwenden wir wiederum Generalsisierte Lineare Gemischte Modelle für propotionalen Antworten der Hörer als abhängiger Variable und *Stimulus* sowie *Gespanntheit* als unabhängige Variablen. Auch hier ergibt sich keine signifikante Beeinflussung der Antworten durch *Gespanntheit* ($\chi^2[1] = 0, 79; n.s.$).

Zusammenfassend lässt sich also sagen:

- Generell unterscheiden sich die Antworten auf die Kontinua *nicht*, es gibt also (ermittelt

über GLMMs) keinen Einfluss der Faktoren *Typ der Verschiebung* oder *Gespanntheit*.

- Nur knapp über die Hälfte der Versuchspersonen weist einen Umkipppunkt auf, kann also den Cue der Grundfrequenz hinreichend zur Kategorisierung der Stimuli in zwei Kategorien nutzen. Doch auch bei den Versuchspersonen mit Umkipppunkt werden die Endpunkte der Kontinua nicht zu 100% als hohe bzw. mittlere Vokale identifiziert.

- Über alle Hörer betrachtet, sind die Steigungen bei *globaler* Grundfrequenzverschiebung etwas flacher.

- Nur die Hörer betrachtet, die den Cue hinreichend nutzen, gibt es aber keine Unterschiede zwischen globaler und lokaler Grundfrequenzverschiebung.

4.4 Diskussion

Wir haben spektral ambige Stimuli, die aus spektralen Kontinua zwischen *bieten* und *beten* und zwischen *bitten* und *betten* ermittelt worden waren, 126 Sprechern präsentiert, wobei die Stimuli in Trägersätze, gesprochen vom selben Sprecher, aus dessen Sprachmaterial die Stimuli resynthetisiert worden waren, eingebettet waren. Wir versuchten, die Vokalkategorisierung der Hörer durch Variation der Grundfrequenz zu beeinflussen, wobei wir erstens *global* und zweitens *lokal* die Grundfrequenz verschoben haben, also über den gesamten Satz hinweg, oder nur in der spektral ambigen Silbe.

Als einer der wichtigsten Befunde dieses Versuchs ist festzuhalten, dass sich die Hörer in jene, die den Cue zur Vokalkategorisierung heranziehen können, und jene, die dies nicht können, unterscheiden lassen. Dies ist indirekt konsistent mit dem Befund aus Hoole und Honda (2011), dass nicht alle, aber manche Sprecher des Deutschen die intrinsische Grundfrequenzeffekte aktiv verstärken, wobei die Autoren dies darauf zurückführen, dass möglicherweise nicht alle Sprecher/Hörer des Deutschen den *intrinsischen* Cue *F1-f0* in gleicher Weise nutzen. Direkt konsistent ist der Befund mit den Daten Traunmüllers, der ebenfalls nur für die Hälfte seiner Versuchspersonen, nämlich Hörern des (eigentlich vokalhöhenreichen) Schwedischen, eine Nutzung des *F1-f0*-Cues nachweisen konnte (vgl. Traunmüller (1991a, 1991b)).

Aus Gründen, die uns selbst nicht klar sind, sind unter Hörern, die die Testmaterialien über Lautsprecher präsentiert bekamen, wesentlich mehr Nicht-Nutzer des Cues zu finden, weshalb sich für diese Gruppe insgesamt keine Kategorisierung in zwei Vokalkategorien feststellen lässt, sondern nur die Tendenz zu einer kontinuierlichen Beeinflussbarkeit, wobei aber die Antworten immer in Richtung der höheren Vokale gebiast bleiben. Wir haben deshalb diese Gruppe aus der hier präsentierten Analyse ausgelassen, und ihre Daten in Anhang B.1.1 präsentiert.

Unter den Hörern, die Lautsprecher benutzt haben, ist die Anteil derer, die den Cue nutzen, größer, wenn auch nur bei knapp über 50%. Interessanterweise nutzen jedoch etwas mehr Hörer den Cue im *ungespannt*-Kontinuum als im *gespannt*-Kontinuum. Auch für die im Anhang präsentierten Daten der Lautsprecherhörer gibt es Hinweise darauf, dass der Cue im *ungespannt*-Kontinuum etwas besser genutzt werden kann. Für die hier gezeigten Kopfhörernutzer lässt sich feststellen, dass die Steigung über alle Hörer hinweg für die *lokale* Grundfrequenzverschiebung etwas höher ist. Dies alles deutet darauf hin, dass es drei Sorten von Hörern gibt: solche, die den Cue der Grundfrequenz zur Vokalkategorisierung nutzen, solche, die ihn nicht nutzen, und einige

wenige, die ihn bedingt nutzen. Diese letztgenannten scheinen ihn bevorzugt nur bei einer größeren Abweichung der Grundfrequenz von der prosodischen *baseline*, also bei lokaler Verschiebung, sowie nur bei *ungespannten* Paaren zu nutzen. Letzteres ist wiederum konsistent mit der Behauptung eines Sonderstatus von ungespannten Vokalen im Deutschen Fischer-Jørgensen (1990); Mooshammer et al. (2001); Hoole et al. (2004); Pape und Mooshammer (2004, 2006a); Hoole (2006); Hoole und Honda (2011), die die Vermutung einschließt, dass dort die Beeinflussung durch den intrinsischen Cue der Grundfrequenz stärker ist.

In der Analyse der Daten aller (Kopfhörer)-Hörer ergibt sich keine signifikante Beeinflussung durch die von uns eingeführten Faktoren *Gespanntheit* und auch nicht durch den *Typ der Verschiebung* (also *lokal* vs. *global*). Dies bestätigt sich v.a. auch für die Daten derjenigen Hörer, die einen Umkipppunkt aufweisen, die also den Grundfrequenz-Cue genutzt haben, um den ambigen Stimulus entweder als hohen oder mittleren Vokal wahrzunehmen – gerade auch bei ihnen sind keine Effekt der zwei Faktoren zu finden. Die Hypothesen 2 und 3, die wir aufgestellt haben, konnten also nicht bestätigt werden.

Dass die Gespanntheitsopposition keinen Einfluss auf die verstärkte perzeptive Nutzung vokalintrinsischer Grundfrequenzvariation bei ungespannten Vokalen im Deutschen habe, zeigten bereits (Pape et al., 2005), obschon die Aufgabe dort etwas entfernter von Sprachperzeption gehalten worden war, denn es mussten dort reine pitch-Unterschiede bewertet werden, also wahrgenommene Tonhöhen. Wir bestätigen dieses Ergebnis mit einem des spekulierten Gebrauchs der intrinsischen Grundfrequenz als *feature enhancer* etwas näherkommenden Experiment. Es bleibt also nur die Möglichkeit bestehen, dass die oben beschriebene stärkere Aktivierung aktiver Prozesse (die den unbestreitbaren biomechanischen Kopplungen zwischen Zunge und Kehlkopf gegenüberstehen) deshalb eingesetzt werden, um die spektral relativ ähnlichen gespannten mittleren und ungespannten hohen Vokale (also /eː-ɪ/) besser voneinander zu unterscheiden (wobei diese Aufgabe im Deutschen natürlich bereits teilweise durch den Längenkontrast übernommen wird), wie Pape und Mooshammer (2006a) vorschlagen.

Zwar untersuchten wir in diesem Experiment nicht das vorgeschlagene /eː-ɪ/-Paar, konnten aber zeigen, dass in der Tat ambige Stimuli mal als mittlerer, mal als hoher Vokal wahrgenommen werden, je nach gegebener Grundfrequenz. Es muss einschränkend darauf hingewiesen werden, dass die Wahrnehmung kontinuierlich, und keineswegs kategorial ausfällt, und dass die Endpunkte der Kontinua sogar bei Versuchspersonen, die einen eindeutigen Umkipppunkt aufweisen, nie die 100% erreichen, die man beispielsweise bei Formantsynthese-Experimenten gewöhnlich erreicht. Durch die Grundfrequenzvariation erhöht sich allerdings die Wahrscheinlichkeit, dass als Antwort entweder der mittlere (bei tieferer *f0*) oder der hohe (bei höherer *f0*) gewählt wird. Hypothese 1 wenigstens kann also bestätigt werden: Trotz des Einflusses *extrinsischer* Faktoren (auf die wir gleich zu sprechen kommen werden) unterstützen *intrinsische* Faktoren, hier also die Grundfrequenz, die Vokalperzeption. Dies unterstützt nur die schon früh in der experimentellen Phonetik (z. B. in Ainsworth (1975)) aufgestellten Vermutungen, dass Hörer sowohl extrinsische als auch intrinsische Cues zur Vokalkategorisierung verwenden. Der Effekt ist vergleichsweise schwach, was aber bei der hier eingesetzten geringen Spannweite der Grundfrequenzvariation (die wesentlich geringer ist als in den Vorgängerstudien, die synthetische Stimuli einsetzten), auch zu erwarten war, zumal, wie erwähnt, in allen hier getesteten Bedingungen auch extrinsische Faktoren eine Rolle spielten.

Einer dieser extrinsischen Faktoren war die durch die globale Grundfrequenzverschiebung bedingte generelle Variation der Vokalhöhe in der gesamten Äußerung. Wir hatten spekuliert,

dass, da sich ja in der gesamten Äußerung der Vokalraum nach oben oder unten verschiebt, das Perzept für den Vokal im Zielwort weiterhin ambig bleiben könnte. Wie sich zeigte, ist dies nicht der Fall, sondern die Grundfrequenzvariation wird trotz der Tatsache, dass auch in den anderen Vokalen eine solche stattfand, zur Wahrnehmung einer sich ändernden Vokalqualität genutzt. Hörer sind also offenbar dazu geneigt, trotz sich gleichzeitig ändernder Rahmenbedingungen die Grundfrequenz in Zweifelsfällen zur Vokalkategorisierung heranzuziehen.

Man kann sich nun fragen, wieso dieser Cue in der *globalen* Grundfrequenzvariation genauso gut genutzt werden kann wie in der *lokalen*. Wahrscheinlich ist dies aber die falsche Frage, und sie sollte eher andersherum gestellt werden: wieso ist die Nutzung des Cues Grundfrequenz zur Vokalhöhenkategorisierung in der *lokalen* Variationsbedingung vergleichbar schlecht wie in der globalen?

Selbstverständlich käme es einem Vergleich zwischen Äpfeln und Birnen gleich, wollte man die Steigungen der Antwortkurven der *lokalen* und *globalen* Grundfrequenzvariation direkt miteinander vergleichen, denn es handelt sich um zwei unterschiedliche Effekte, die die Nutzung der Grundfrequenz als Vokalhöhencue stören: die vermutete generelle Anhebung bzw. Absenkung der Vokalhöhe aller Vokale durch die globale Grundfrequenzvariation auf der einen, und die intonatorische Funktion der Grundfrequenz auf der anderen Seite. Die lokale Variation kann – wie die prosodische Labelung, die wir anfertigen ließen, zeigt – zu mindestens drei wahrnehmbaren intonatorischen Kategorien führen. Es ist sehr wahrscheinlich, dass auch die Versuchspersonen einen Teil der Grundfrequenzvariation als prosodische Variierung wahrgenommen haben. Fowler und Brown (1997) vermuteten hinter den *intrinsic pitch*-Effekten (siehe Einleitung), die sie fanden, eine perzeptuelle Kompensation für die Effekte der verschiedenen Vokalhöhen auf die Grundfrequenz, um die intonatorische Struktur nicht zu stören[2]; es gebe also ein *parsing* der Grundfrequenzinformation in intrinsische (segmentale) und intonatorische (suprasegmentale) Anteile; allerdings muss dazugesagt werden, dass diese Kompensation bestenfalls partiell ist, da die Größenordnungen für den intrinsic pitch, die sie finden, nur ein Zehntel der vokalintrinsischen Grundfrequenz betragen. Allerdings fanden Pape und Mooshammer (2006b, 2008) sprachspezifische Unterschiede in diesem *parsing*, dass nur ausgeprägt zu sein scheint, wenn die Sprache eine reiche Vokalhöhendifferenzierung aufweist. Das Ausmaß der Kompensation und des *parsings* in zwei Informationsarten ist also variabel. Dies führte uns zu dem Gedanken, dass das *parsing* auch innerhalb einer Sprache variabel sein könnte, je dringender der Grundfrequenz-Cue zur Adjustierung der Bestimmung der Vokalhöhe benötigt würde - so z. B. in Zweifelsfällen wie den in diesem Experiment gegebenen ambigen Stimuli. Offenbar wird der Grundfrequenz-Cue aber nur in geringem Ausmaß als intrinsisches Vokalhöhenmerkmal genutzt, obschon der betroffene Vokal spektral ambig ist, denn die Antwortkurven sind für diesen Teil des Experiments ebenfalls sehr flach.

Es gibt Alternativen zur Erklärung des *intrinsic pitch*-Effektes als Resultat einer Kompensation in Form der Hypothese der virtuellen Tonhöhen, also der Wahrnehmung der Tonhöhe als Resultat einer Beeinflussung der Tonhöhe durch das Spektrum eines komplexen Signals (Stoll, 1984). Dieser Erklärungsansatz war u. a. in Pape (2005) und Pape et al. (2005), wo unter anderem gerundete und ungerundete Vokale gleicher Vokalhöhe als Stimuli benutzt worden waren, abgelehnt worden (vgl. aber auch, neben Stoll (1984), die kurze Diskussion hierzu in Traunmüller (2005)). Es ist nicht ausgeschlossen, dass die gleichbleibende Einhüllende des Filterspektrums der

[2]Siehe Thorsen (1984) für eine Beschreibung der innerhalb eines Sprechers als invariant angenommenen intonatorischen Muster und die Veränderung dieser Muster an der Oberfläche durch die Einflüsse der intrinsischen Grundfrequenz

Stimuli die *intrinsic pitch*-Wahrnehmung dergestalt beeinflusst hat, dass für die Hörer kein Grund bestand, in größerem Ausmaß ein *parsing* der Grundfrequenzinformation vorzunehmen, so dass der größte Teil davon als Intonation wahrgenommen wurde. Dies muss an dieser Stelle jedoch Spekulation bleiben, da dies über das Ziel dieser kleinen Untersuchung im Rahmen dieser Dissertation hinausschießt. Der Alternativansatz könnte aber vielleicht auch erklären, weshalb es in der Mitte der Kontinua zu einem – bislang nicht diskutiertem – etwas paradoxem Antwortverhalten kommt, wenn über einen kleinen Bereich der Kontinua die Wahrscheinlichkeit für *hohe Vokalhöhe*-Antworten sinkt, obwohl die Grundfrequenz steigt. Es bleibt aber festzuhalten, dass auch in diesem Teil der Untersuchung grundsätzlich ein das Vokalhöhenperzept beeinflussender Effekt vorhanden ist - ein wenig *parsing* in intonatorische und vokalhöhendefinierende Information muss also durchaus stattgefunden haben, wenn auch nicht in dem erwarteten Umfang.

Wir haben leider keine Vergleichsdaten, die z. B. aus den gleichen Vokaltokens, wie sie hier verwendet wurden, die man aber ohne Trägersatz und sonstigen Kontext präsentieren müsste, entnommen werden könnten, um abzuschätzen, ob die kontextuellen Einflüsse, also die eher als extrinsische Vokalraumverschiebung zu deutende globale Verschiebung der Grundfrequenz, oder die als Intonation zu deutende lokale Verschiebung, die Antwortkurven so abflachten, oder ob die rein intrinsisch zu bezeichnende Information in isoliert präsentierten Stimuli die Kurven wirklich steiler gemacht hätten. Auch dies muss leider zukünftigen Arbeiten überlassen werden.

Es bleibt also als wichtigste Erkenntnisse dieses Experiments festzuhalten:

- Es gibt trotz konfligierender extrinsischer Information und trotz der intonatorischen Funktion der Grundfrequenz auch in fließender Rede einen (schwach ausgeprägten) *intrinsischen* Einfluss auf das Vokalhöhenperzept deutscher Hörer
- Nur ein Teil der Hörer scheint sensitiv für den vokalhöhendefinierenden (bzw. -verstärkenden) Einfluss der Grundfrequenz zu sein
- Es besteht nur sehr schwache Evidenz dafür, dass Versuchspersonen, den den Cue nur bedingt nutzen, ihn bevorzugt in ungespannt-Vokalkontinua nutzen. Wesentlich verlässlicher sind die Daten, die aussagen, dass, wenn Hörer den Cue Grundfrequenz zur Vokalhöhenbestimmung nutzen, sie ihn in ungespannten wie gespannten Vokalkontinua in gleichem Ausmaß nutzen.

Kapitel 5

Fazit und Ausblick

Fassen wir die wichtigsten Erkenntnisse dieser Arbeit noch einmal zusammen. Aus den Literaturüberblicken in den verschiedenen Einleitungskapiteln dieser Arbeit lernten wir[1]:

- Jeder Sprecher hat eine individuelle Stimme, mit individuell verschiedenen Formantlagen, mittlerer Grundfrequenz, Phonationsverhalten usw.; die Stimme kann der Sprecher aktiv modulieren und den Notwendigkeiten sprachlicher Kommunikation anpassen; so kann er z. B. den *vocal effort* erhöhen, wenn der Hörer weit entfernt ist. Auch diese Modulationen haben Auswirkungen auf die akustischen Merkmale des Sprachsignals. Davon zu trennen ist die linguistisch bedingte Variation bestimmter Parameter (vgl. 1).

- Hörer sind offenbar in der Lage, sich in eine Sprecherstimme einzuhören und die individuelle Variation akustischer Korrelate der Sprachproduktion zu berücksichtigen, um aus dieser Variation abzuleiten, wie die akustischen Konsequenzen der Artikulation zu erwarten sind. Abweichungen von dieser Erwartung – wie Sprecherwechsel oder stärker ausgeprägte Variation innerhalb eines Sprechers aus extralinguistischen Gründen – beeinflussen die Perzeption linguistisch relevanter Information und verlangen nach einer perzeptuellen Adjustierung (vgl. 1.1, 1.2, 1.6).

- Trotz der interindividuellen Unterschiede gibt es statistische Wahrscheinlichkeiten und Korrelationen bezüglich des Verhältnisses bestimmter akustischer Korrelate der Sprachproduktion. In der Regel haben Männer die größeren Kehlköpfe, und damit die tieferen Grundfrequenzen, und gleichzeitig die größeren Ansatzrohre, und damit tiefere Formantlagen, als es Frauen haben, und deren Werte sind tiefer als jene von Kindern. Zwischen den genannten drei Gruppen ist dieser Zusammenhang sehr deutlich ausgeprägt, innerhalb der Gruppen weniger (z.B. die relativ schwachen Korrelationen zwischen Körpergröße oder -gewicht und der Grundfrequenz bzw. den Formanten); dennoch kann man auch innerhalb der Gruppen von einem gewissen statistischen Zusammenhang zwischen diesen Maßen ausgehen. In anderen Worten: es besteht ein relativ zu verallgemeinender Zusammenhang zwischen dem Alters- (Kind vs. Erwachsener) bzw. Geschlechtsdimorphismus und den hiermit kovariierenden Kehlkopf- und Ansatzrohrgrößen (-längen). Wird hiervon abgewichen, sinken Erkennungsraten (vgl. 1.1, 1.2, 1.3).

[1] Da wir hier nochmals eine Kurzüberblick bieten, verzichten wir auf Quellenangaben; diese sind in den entsprechenden Kapiteln zu finden.

- Letzteres ist wohl auch damit zu erklären, dass offenbar eine grobe, aber dafür ohne Involvierung höherer Verarbeitungsbereiche funktionierende Normalisierung für diese alters- und geschlechtsbedingten Kovariationen bereits durch die Aufteilung der Basilarmembran im menschlichen Ohr in kritische Bänder vorgegeben ist. Die mittleren Grundfrequenzen und die linguistisch nur wenig relevanten und daher relativ invarianten höheren Formanten ab *F3* liegen zwischen Kindern, Frauen und Männern je circa ein kritisches Band auseinander. Wegen dieser „festen Verdrahtung“ ist dieser Vorgang sehr schnell, und deshalb auch auf kurze vokoide Signalanteile anwendbar, aber auch nur sehr grob. Auf diese Verhältnisse basierende algorithmische *intrinsische* Normalisierungsmethoden zeichnen sich analog hierzu auch eher dadurch aus, dass sie die groben Unterschiede zwischen Männern, Frauen, und Kindern hinwegnehmen, aber doch auch viel der individuellen Variation übriglassen. Für eine weitergehende Normalisierung sind offenbar gerade auch für menschliche Hörer kontextuelle Informationen notwendig, die in höheren Bereichen verarbeitet werden (vgl. 1.3, 1.6, 1.7, 4.1).

- Ein Sprecher ist (fast) immer auch ein Hörer seiner eigenen Sprachproduktion. Es steht also, neben dem somatosensorischen Feedback, immer auch auditorisches Feedback zur Verfügung. Dieses spielt im Spracherwerb eine große Rolle, verliert dann aber an Wichtigkeit, wird aber wieder wichtig, wenn Störungen eintreten. Man kann vermuten, dass Sprecher Repräsentationen von auditorischen Zielregionen, die sie erreichen wollen, zur Feinjustierungen der Produktion zur Aufhebung der akustischen Konsequenzen dieser Störungen oder *Perturbationen* nutzen, wenn dies nötig wird. Man hat gezeigt, dass die erlernte, schnellere, und daher normalerweise bevorzugte Feedforward-Kontrolle bei Perturbationseinfluss recht schnell rekalibriert werden kann, was sich auch daran zeigt, dass die Ersetzung eines perturbierten akustischen Feedbacks durch Rauschen sich nicht etwa dadurch äußert, dass Sprecher wieder ihre ursprüngliche Konfiguration herstellen, sondern weiterhin wie unter Perturbation produzieren.

  Somatosensorisches Feedback ist schneller und wohl auch genauer als die auditorische Feedback-Schleife, so dass auf mechanische Perturbationen in der Regel schneller und vollständiger kompensiert werden kann als auf Störungen im auditiven Kanal (vgl. 3.1).

- Zumeist ist die artikulatorische Variation – ob sie nun zu einem linguistisch relevanten oder zu einem extralinguistisch relevanten Effekt führen soll – durch die Beteiligung mehrerer Artikulatoren gekennzeichnet. Diese Kovariation von Artikulatoren lässt sich danach unterteilen, ob die Artikulatoren voneinander abhängig sind, also schlicht durch biomechanische oder akustische Kopplung nicht unabhängig voneinander bewegt werden können, oder ob die Artikulatoren voneinander unabhängig agieren können. Außerdem muss man unterscheiden zwischen merkmalsdefinierenden und merkmalsverstärkenden Gesten. Diese beiden Unterscheidungen sind aber nicht immer leicht getroffen. So ist bei der intrinsischen Grundfrequenz umstritten, ob man es hier mit voneinander abhängigen Artikulationen (wobei die zweite Unterscheidung in merkmalsdefinierend und -verstärkend – zumindest weitgehend – entfällt), oder aber mit unabhängigen Artikulationen (wobei man die Zungenhöhenvariation dann noch als merkmalsdefinierend, und die Grundfrequenzvariation als merkmalsverstärkend bezeichnen müsste) zu tun hat, oder ob man einen mittleren Weg geht, und eine merkmalsverstärkende unabhängige Grundfrequenzvariation nur in manchen Fällen für gegeben hält (vgl. 1.5).

Bei der (extralinguistischen) *vocal effort*-Variation ist es noch unklarer, welche der zahlreichen beteiligten Variablen unabhängig oder abhängig sein soll. Immerhin lässt sich sagen, dass zwei der Variablen, die laryngalen Maße und der Öffnungsgrad des Ansatzrohrs, wahrscheinlich nicht unabhängig voneinander sind, was möglicherweise eine gröbere Abweichung des vokalhöhendefinierenden *F1-f0*-Abstandes, auf den wir noch zu sprechen kommen, verhindern soll, denn es lässt sich feststellen, dass bei *vocal effort*-Variation *f0* und *F1* kovariieren (vgl. 1.7, 2.2.1).

- Für koartikulationbedingte Variation kann man zeigen, dass oft die *merkmalsverstärkenden* Gesten weniger betroffen sind als die *merkmalsdefinierenden* Gesten (vgl. 1.5). In mechanischen Perturbationsexperimenten werden in der Regel die merkmalsdefinierenden Gesten eingeschränkt, und man weiß, dass für diese Perturbation durch einen verstärkten Einsatz der (normalerweise merkmalsverstärkenden) Gesten kompensiert werden kann (vgl. 3.1). Es liegt nahe, zu vermuten, dass auch für Perturbationen des auditorischen Feedbacks mit kompensatorischen Antworten in mehreren Parametern zu rechnen sein wird, zumal in diesem Fall damit zu rechnen ist, dass der für eine akustische Perturbation kompensierende Sprecher konfligierendem Feedback aus der somatosensorischen und der auditorischen Schleife ausgesetzt sein wird, was die vollständige Kompensation über eine merkmalsdefinierende Geste stark einschränken dürfte (vgl. 2.2.1).

- Die bereits erwähnte vokalintrinsische Grundfrequenz zeichnet sich dadurch aus, dass hierdurch eine positive Korrelation zwischen Zungenhöhe und Rate der Stimmlippenschwingung, und daraus resultierend eine negative Korrelation zwischen *F1* und *f0* zu verzeichnen ist. Vokalhöhenvariation ist also durch eine gegenläufige Bewegung von *F1* und *f0* gekennzeichnet (vgl. 1.5).

  Geht man von einer aktiven oder zumindest in manchen Fällen aktiv einen automatischen Effekt verstärkenden Steuerung der intrinsischen Grundfrequenz aus, muss man auch davon ausgehen, dass der am stärksten mit Vokalhöhe korrelierte Parameter *F1 in Relation zur Grundfrequenz* wahrgenommen wird, also dass nicht der *F1*-Wert an sich, sondern der Abstand zwischen *F1* und *f0* das bessere, auditorisch entscheidendere Vokalhöhenkorrelat ist (vgl. 1.5). In der Tat wurde festgestellt, dass *F1-f0*[Bark] das bessere Vokalhöhenkorrelat ist, wenn auch mit etlichen Einschränkungen. Eine dieser Einschränkungen ist die, dass es eine Inter-Hörer-Variabilität in der Nutzung dieses Cues gibt – einige nutzen ihn, einige eher nur *F1* –, und weiters, dass es umso mehr Nicht-Nutzer des Cues gibt, desto weniger Vokalhöhenunterschiede die untersuchte Sprache aufweist. Interessanterweise verdichten sich in den letzten Jahren die Hinweise darauf, dass auch die aktive Nutzung/Verstärkung der vokalintrinsischen Grundfrequenz – zumindest in manchen Distinktionen – von dem Vokalhöhenreichtum der untersuchten Sprache abhängt. Ebenso wie in den Perzeptionsexperimenten, die selbst für vokalhöhenreiche Sprachen aufzeigen, dass nicht alle Versuchspersonen gleich sensitiv für den *F1-f0*-Cue sind, ist auch für den *aktiven* Gebrauch der intrinsischen Grundfrequenz festzustellen, dass nicht alle Versuchspersonen diese Möglichkeit nutzen (vgl. 1.7, 1.5, 4.1).

Wie gesagt, spielt auditorisches Feedback eine große Rolle während des Spracherwerbs, und verliert später an Wichtigkeit. Man weiß aber, dass das auditorische Feedback relativ lange – bis zum frühen Erwachsenenalter hinein – seine wichtige Rolle beibehält. Dies hat mit einiger

Sicherheit damit zu tun, dass während der Pubertät ein zwar geschlechtsspezifisch unterschiedlich ausgeprägtes, aber bei beiden Geschlechtern deutliches Wachstum der Artikulatoren einsetzt, dass letztendlich in dem stark ausgeprägten Dimorphismus, den wir oben beschrieben haben, resultiert. Die bereits erlernte Feedforward-Kontrolle muss während dieser Zeit ständig nachjustiert werden. In gewisser Weise ist also die Mutation eine zwar relativ langsam fortschreitende, aber sehr global ausgeprägte Perturbation, für die ständig kompensiert werden muss. Da die Änderungen, die sowohl den Kehlkopf und das Ansatzrohr betreffen, nicht immer gleichzeitig stattfinden, ist auch mit einer Perturbation der Vokalhöhe durch eine Art von Mismatch von Grundfrequenz und Formantlagen zu rechnen.

Doch auch nach erfolgter Mutation gehen ständig weitere körperliche Veränderungen vor sich, die auch in Veränderungen der Artikulatoren resultieren. Wir wollten im ersten experimentellen Kapitel dieser Dissertation (2) die akustischen Resultate dieser ständig langsam fortschreitenden körperlichen Auswirkungen messen, und bestimmen, ob sich hierbei kompensatorisches Verhalten finden lässt.

- Wir führten Messungen der Grundfrequenz und der ersten drei Formanten durch. Wir fanden in aus verschiedenen Lebensjahrzehnten stammenden Aufnahmen mehrerer Sprecher und Sprecherinnen anhand von Messungen in Schwa-Vokalen und von als äquivalent festgestellten Messungen in allen stimmhaften Signalanteilen ähnliche altersbedingte Variationen, wie sie auch oft in der Literatur beschrieben wurden, nämlich Änderungen, die besonders prominent die Grundfrequenz und den ersten Formanten betreffen. Für alle Sprecher und für alle unterschiedlichen Messungen, die wir durchführten, wurden für die Formanten 2 und 3 nur unsystematische und/oder insignifikante Variationen gefunden. Für Frauen wurde ein Absinken der Grundfrequenz festgestellt, und für Männer zunächst ein Absinken, in späterem Alter ein Anstieg der Grundfrequenz. In den meisten Fällen – zu den Abweichungen kehren wir später zurück – zeigte der erste Formant ein ähnliches Muster. Tatsächlich offenbarte die Untersuchung von Aufnahmen, die in kurzen Abständen über mehrere Jahrzehnte mit den gleichen Sprechern – einer Frau und einem Mann – gemacht wurden, dass langfristig gesehen Grundfrequenz und erster Formant kovariieren - bei der Frau sinken beide Parameter in in etwa der gleichen Rate, beim Mann sinken beide Parameter zunächst, um nach einem in beiden Parametern zur etwa gleichen Zeit zu findendem Punkt wieder anzusteigen. Auch bei dieser Untersuchung variierten *F2* und *F3* etwas, allerdings unsystematisch, und änderten sich dementsprechend über die Jahrzehnte betrachtet nicht, mit Ausnahme des zweiten Formanten bei der Königin, der leicht sinkt.

- Nachdem wir ausschließen konnten, dass diese Kovariation eine Artefakt der Messung sein könnte, diskutierten wir mögliche Gründe dieser Kovariation von *f0* und *F1*. Wie die altersbedingte Kovariation von *f0* und *F1* bei gleichzeitig wenig Variation in *F2* und praktisch keiner Variation in *F3* zeigt, erinnert dieses Muster an Variation erstaunlicherweise eher an die akustischen Effekte, die Traunmüller für *vocal effort*-Variation findet (siehe Abbildung 1.5 auf Seite 19), als an die Variation zwischen Kindern und Erwachsenen, wo eine Absenkung sowohl der Grundfrequenz als auch aller Formanten zu finden ist. Es wäre interessant, zu prüfen, ob mit den hier gemessenen *f0*- und Formantwerten synthetisierte Äußerungen - die weitere akustische Cues auf das Sprecheralter nicht enthalten dürften - eher zur Perzeption von Alters-, oder eher zu *vocal effort*-Variation führen würden. Grundsätzlich jedoch halten wir es nicht für wahrscheinlich, dass die hier aufgeführten akustischen Daten ei-

ne Art von *vocal effort*-Variation abbilden, obschon es nicht ganz auszuschließen ist, dass eine *vocal effort*-Erhöhung für den Wiederanstieg beider Parameter im hohen Alter eine Rolle spielt, da bekannt ist, dass die Effizienz der Phonation im höheren Alter nachlässt; auch altersbedingter Hörverlust könnte eine Rolle spielen und den Sprecher den *vocal effort* erhöhen lassen. Wir können an dieser Stelle aber bestenfalls spekulieren.

- Eines aber haben die hier präsentierten Daten und die akustischen Konsequenzen einer *vocal effort*-Variation gemein: ein Sprecher, der *vocal effort* variiert, tut dies u. a. über Variation der Grundfrequenz und des Öffnungsgrads des Kiefers - sein Ansatzrohr bleibt selbstverständlich identisch. Unsere Daten suggerieren, dass sich auch bei den hier untersuchten alternden Sprechern das Ansatzrohr an sich sich nicht über die Jahrzehnte wesentlich geändert hat, denn sonst müsste man auch eine altersbedingte Variation v.a. auch in *F3* feststellen. Im Gegensatz zur altersbedingten Kovariation von Grundfrequenz und *allen* Formanten beim Wandel vom Kind zum Erwachsenen ist hier nur für den ersten Formanten eine konsistente und signifikante Variation feststellbar. Auch der Überblick über die physiologischen Daten zu den Abmessungen des Ansatzrohrs konnte keine eindeutige Vokaltraktverlängerung mit dem Alter aufzeigen. Insofern stehen unsere Daten in Widerspruch zu Modellierungen der Altersstimme, die von einer generellen Vokaltraktverlängerung ausgehen, so wie beispielsweise Linvilles Modell (siehe Abbildung 2.2 auf Seite 34). Da wir keine Aussagen über nicht-neutrale Vokale abgeben, bleiben Modellierungen, die von altersbedingten Veränderungen der Artikulation und damit verbundener Verringerung der Periphizität des Vokalraums ausgehen, von unserer Behauptung unberührt (vgl. 2.1).

  Relativ eindeutig sind hingegen die physiologischen Beschreibungen der altersbedingten Änderungen des Kehlkopfs allgemein und der Stimmlippen und ihrer Masse im besonderen. Diese Änderungen können sehr gut die hier und anderswo beschriebenen Muster für die altersbedingte Variation der mittleren Grundfrequenz - im wesentlichen ein Absinken bei Frauen und ein eher V- oder U-förmiger Verlauf bei Männern – erklären (vgl. 2.1).

- Die von uns gemessenen altersbedingten Veränderungen der mittleren Grundfrequenz sind um Größenordnungen stärker als die Variation der intrinsische Grundfrequenz zu einem Zeitpunkt und im gleichen – auch prosodischen – Kontext. Dies macht es für uns vorstellbar, dass die Veränderung der Grundfrequenz als eine Perturbation des Vokalhöhenperzeptes wahrgenommen wird – wofür die Sprecher dann kompensieren. Da die Grundfrequenz bestenfalls ein merkmalsverstärkendes Element der Vokalhöhe ist, werden sie nicht die Grundfrequenzproduktion verändern, sondern das merkmalsdefinierende akustische Korrelat - den ersten Formanten. Modellierungen der artikulatorisch-akustischen Zusammenhänge, aber auch physiologische Daten zum Formanttuning bei professionellen Sängerinnen legen nahe, das am wahrscheinlichsten hierbei eine Variierung der Kieferöffnung eine Rolle spielen dürfte.

  Die Hypothese ist also, dass eine generelle altersbedingte *F1*-Veränderung nicht physiologisch bedingt ist, sondern aus phonetischen Gründen – zur Beibehaltung der Vokalhöhe – aktiv gesteuert wird (2.3).

Um die Plausibilität dieser Hypothese zu testen, wurden in der Folge Perturbationsexperimente (in Kapitel 3) mit künstlich verändertem auditorischem Feedback durchgeführt.

- Der direkteste Test ist die Perturbation der Grundfrequenz. Die Nutzbarkeit eines *f0*-Perturbationsexperiments zur Überprüfung eines Einflusses der Grundfrequenz auf die Vokalhöhenwahrnehmung – die sich durch eine kompensatorische Variation des ersten Formanten zeigen sollte – ist allerdings dadurch eingeschränkt, dass auch und vor allem in dem perturbierten Parameter kompensiert wird, wobei dieser Vorgang offenbar so automatisch verläuft, dass man ihn auch nicht verhindern kann, wenn man die Versuchspersonen ausdrücklich dazu auffordert, *nicht* zu kompensieren (vgl. 3.1). Wir mussten also das Vorhaben um eine direktere Beeinflussung der Vokalhöhe ergänzen. Dies geschah durch eine Perturbation des ersten Formanten in einem gesonderten Experiment, und der Messung der unter Perturbation produzierten *F1*- und *f0*-Werte (3.2).

- Zunächst scheint es nicht sehr zielführend zu sein, den ersten Formanten zu perturbieren, denn hierfür wird in der Hauptsache mit einer Kompensation in *F1* zu rechnen sein. Auch mit einer gewissen negativen Korrelation der produzierten *F1*-Werte mit den *f0*-Werten musste wegen des Automatismus der intrinsischen Grundfrequenz gerechnet werden - wenn die Kompensation in *F1* über die Veränderung der Zungenhöhe produziert wird, ist wegen der biomechanischen Kopplung mit einem gewissen Einfluss auf die Grundfrequenzwerte zu rechnen. Dieser Automatismus bedingt, dass die Grundfrequenzwerte umso mehr von der *baseline*, also den produzierten Werten in den Experimentphasen ohne Perturbation, abweicht, desto mehr Kompensation im Parameter *F1* auftritt. Ein solcher Effekt ist auch in früheren Studien bereits gefunden worden (vgl. Abbildung A.3 auf Seite 197).

- Wir machten uns zunutze, dass bekannt ist, dass Kompensation für akustische Perturbationen auf ein gewisses Maß beschränkt sind, die Zungenhöhenproduktion und damit die *F1*-Produktion also ab einer gewissen Perturbationsstärke – vermutlich wegen des Einflusses des konfligierenden somatosensorischen Feedbacks – geblockt ist. Sollte es der Fall sein, dass Sprecher/Hörer des hier untersuchten Deutschen *F1 in Relation zu f0* als Vokalhöhenkorrelat perzibieren, könnten sie die Grundfrequenz *aktiv* variieren, um dem Erreichen eines zufriedenstellenden auditorischen Vokalhöhenperzepts näher zu kommen.

- In der Tat zeigen die Ergebnisse, konsistent mit Vorgängerstudien, über alle Sprecher eine kompensatorische Gegenbewegung des produzierten *F1*-Wertes, und eine Verschiebung der produzierten *f0*-Werte in Gegenrichtung hierzu, wenn die Sprecher einer Perturbation ausgesetzt waren. Dies alleine ist, wie gesagt, auch über die biomechanische Kopplung der Zunge und des Kehlkopfes erklärlich, zumal die Änderung in *f0* gegenüber derjenigen in *F1* gering ausfällt. Ein zweiter Befund ist der, dass sich die produzierten *F1*-Werte in zwei Perturbationsstärken nicht unterscheiden, eine weitergehende kompensatorische *F1*-Verschiebung also geblockt zu sein scheint. In der Tat zeigen aber einige Sprecher in diesem Fall einen verstärkten Einsatz der Grundfrequenz – was, wie erwähnt, als indirekter Hinweis verstanden werden kann, dass diese Sprecher als Hörer zur Vokalhöhenperzeption auch die Grundfrequenz benützen.

Daraufhin wurde der Einfluss eine Grundfrequenzverschiebung auf die Vokalproduktion geprüft. Hierzu wurde – trotz der erwähnten Einschränkungen dieser Methode – eine Quasi-Echtzeit-Grundfrequenzperturbation auf das auditorische Feedback der Sprecher ausgeübt (3.3).

- Eine weitere Einschränkung der Aussagekraft dieses Experiments bestand darin, dass die Daten von sechs Sprecherinnen offensichtlich durch Artefakte des Formantextraktionsalgo-

rithmus gestört waren und deshalb ausgeschlossen werden mussten. So sehr das Auffinden dieses offensichtlichen Artefaktes ein Nebenbefund sein mag, ist dies doch für weitere Untersuchungen des *F1-f0*-Zusammenhanges ein nicht zu vernachlässigender Befund und verlangt nach größerer Vorsicht. Da sowohl im *F1-* als auch in diesem *f0*-Perturbationsexperiment die gleichen Sprecher und die gleichen Messmethoden verwendet wurden, ist damit zu rechnen, dass auch die Daten des *F1*-Experiments von diesem störenden Einfluss beeinträchtigt waren, ohne das dies dort besonders aufgefallen wäre.

- Da eine gewisse Korrelation zwischen vertikaler Kehlkopfposition und Grundfrequenz bekannt ist, maßen wir in diesem Experiment neben der Grundfrequenz die ersten drei Formanten. In der Tat lassen sich in den Formanten nur leichte Produktionsänderungen feststellen. *F2* zeigt nur eine Tendenz dazu, mit *f0* anzusteigen, *F3* jedoch ist deutlicher mit der Grundfrequenz korreliert. Wir deuten dies als durch Veränderung der vertikalen Kehlkopfposition verursachter Effekt einer Verkürzung/Verlängerung des Ansatzrohrs. Die Ansatzrohrlänge scheint hier also leicht positiv mit *f0* zu korrelieren. Umso höher ist der Effekt auf *F1* einzuschätzen: *F1* wird in Gegenrichtung zur *f0*-Produktion realisiert, so, wie man es für eine durch unvollständige Kompensation in *f0* verschobene Vokalhöhenwahrnehmung und dadurch ausgelöste Vokalhöhenkompensation erwarten würde. Wir halten diesen Effekt für ein Beispiel einer doppelten Perturbation: Im Experiment wird *f0* perturbiert. Für diejenigen Sprecher, die für die Perturbation nicht vollständig (oder auch gar nicht) kompensieren, erscheint hierdurch die Grundfrequenz nach wie vor in Perturbationsrichtung verschoben. Für diejenigen Sprecher/Hörer, die die Vokalhöhe mit Bezug auf die Grundfrequenz bewerten, erscheint nun die Vokalhöhe beeinträchtigt – wofür sie kompensieren.

  Wir finden diesen Effekt allerdings auch in diesem Experiment nur in einigen Versuchspersonen. Da teilweise auch relativ deutlich und teilweise sogar vollständig mit *f0* für die *f0*-Perturbation kompensiert wird, war damit allerdings auch zu rechnen – selbst wenn diese Versuchspersonen den *F1-f0*-Cue nutzen sollten, sollten sie wegen des nur geringen Unterschiedes keinen Grund haben, eine Vokalhöhenperturbation wahrzunehmen.

Die Schlussfolgerung aus beiden Perturbationsexperimenten ist also, dass die Grundfrequenz aktiv eingesetzt werden *kann*, um die Vokalhöhe zu beeinflussen, was nur dann der Fall sein sollte, wenn die Sprecher auch als Hörer die Grundfrequenz in ihrer Vokalhöhenperzeption mit berücksichtigen, und dass Vokalhöhenregulationen durch Grundfrequenzperturbation ausgelöst werden *können*. Genau den letzten Effekt vermuteten wir als Grund für die Kovariation von *f0* und *F1* bei alternden erwachsenen Sprechern. Natürlich schließen wir aber nur aus den Produktionsdaten auf die vermutliche perzeptuelle Nutzung des *F1-f0*-Cues. Ein direkterer Test ist in diesem Fall natürlich ein Perzeptionsexperiment (4).

- Wir präsentierten spektral ambige Stimuli mit verschiedenen Grundfrequenzen, um den Einfluss der Grundfrequenzvariation auf die Vokalkategorisierung zu bestimmen. Wir verwendeten Resynthesen natürlicher Sprache, um zu vermeiden, dass es Effekte gibt, die auf die Rekalibrierung der Perzeption auf einen anderen „Sprecher“ zurückzuführen gewesen wären.

- In den letzten Jahren haben sich die Hinweise verdichtet, dass es eine Besonderheit des Deutschen ist, dass für ungespannte Vokale die intrinsische Grundfrequenz aktiver eingesetzt wird als in gespannten, wo der intrinsische-Grundfrequenz-Effekt eher das Resultat

der biomechanischen Kopplung zwischen Zunge und Kehlkopf ist. Dies lässt vermuten, dass *f0* für ungespannte Vokale des Deutschen *merkmalsverstärkend* eingesetzt wird, woraus sich schlussfolgern ließe, dass eventuell Hörer des Deutschen bei ungespannten Vokalen sensitiver auf Grundfrequenz-Cues zur Vokalhöhe reagieren als bei gespannten Vokalen (vgl. 1.5).

Deshalb untersuchten wir ein Kontinuum zwischen ungespannten Vokalen, und ein Kontinuum zwischen gespannten Vokalen. Es ließen sich keine Unterschiede in der Nutzung der Grundfrequenz als Cue zur Vokalhöhe feststellen.

- Grundsätzlich ist die Nutzung des Cues sehr kontinuierlich, und führt offenbar nicht zu kategorialen Perzepten, selbst nicht an den Endpunkten. Konsistent mit Vorgängerstudien (vgl. 1.7) ist festzustellen, dass nur ein Teil der Versuchspersonen (circa 50%) die Grundfrequenz zur Vokalkategorisierung so nutzen kann, so dass beide Enden des Kontinuum zu mehr als Zufallsniveau als eine der beiden Vokalhöhen identifiziert wird. Auch hier ist kein Vorteil auf Seiten des ungespannt-Kontinuums zu erkennen, denn in beiden Kontinua verwendet nur etwa die Hälfte der Versuchpersonen die Grundfrequenz als Cue zur Vokalhöhe.

- Zur schlechten Nutzbarkeit des Cues hat sicher beigetragen, dass wir das Kontinuum an jeweils akzentuierter Position in einen Trägersatz eingebaut haben. In einem Fall haben wir die Grundfrequenzverschiebung *global* über die gesamte Äußerung vorgenommen, und damit einen *extrinsischen* Cue über die zu erwartende Vokalhöhe geliefert (vgl. 1.5, 4.1), im anderen Fall haben wir die Grundfrequenzvariation *lokal* nur in der in Frage stehenden Silbe vorgenommen, und damit eine intonatorische Variation eingeführt (vgl. 4.1). Beide Variationen führen offenbar zu einem ähnlichen Ausmaß an Einschränkung für die Nutzbarkeit des *intrinsischen* Cues. Leider haben wir verabsäumt, zum Vergleich die gleichen Kontinua in Isolation zu testen, um abschätzen zu können, inwieweit die extrinsische und intonatorische Variation wirklich die Nutzbarkeit einschränkt, und ob in Isolation präsentierte Kontinua tatsächlich besser hätten identifiziert werden können. Jedenfalls ist trotz der extrinsischen Einflüsse und der Notwendigkeit, die Grundfrequenz in ihre vokalintrinsische und intonatorische Funktion zu parsen, ein Einfluss der Grundfrequenz auf die Vokalhöhenperzeption nicht abzustreiten.

Grundsätzlich haben wir also gezeigt, dass Sprecher/Hörer des Deutschen teilweise den *F1-f0*-Cue zur Vokalhöhenperzeption nutzen, und vermutlich deswegen bei Blockade einer kompensatorischen *F1*-Produktionsverschiebung unter vokalhöhenbeeinflussender *F1*-Perturbation unabhängig von *F1* die Grundfrequenz einsetzen, und dass manche Hörer/Sprecher des Deutschen auch bei einer reinen *f0*-Perturbation überraschenderweise Kompensationen der Vokalhöhe vornehmen. Wir müssen streng genommen vorsichtig sein, deswegen darauf zu schließen, dass wir damit gezeigt haben, wie plausibel die Hypothese zu der altersbedingten *F1-f0*-Kovariation sei, denn, wie der Überblick in der allgemeinen Einleitung (1) zeigte, wird der Cue *F1-f0* nicht nur in Sprachen, die sich im Reichtum an Vokalhöhen unterscheidet, variabel benutzt, sondern auch in eng verwandten Sprachen mit ähnlicher Vokalhöhendistinktion, und selbst innerhalb einer Sprache scheint es Unterschiede in der Nutzung des Cues zu geben, abhängig davon, ob man Vokalhöhe bei vorderen oder hinteren, gerundeten oder ungerundeten Vokalen betrachtet. Es ist auch nicht immer ein linearer Zusammenhang zwischen der Anzahl der Vokalhöhenunterscheidungen und der Nutzbarkeit des *F1-f0*-Cues zu ziehen, wie die gute Nutzbarkeit des Cues im eher vokalhöhenarmen Französischen verglichen mit der doch eher offenbar eingeschränkten Nutzbarkeit

und Beschränkung auf eher binäre Unterteilungen im Amerikanischen Englisch zeigt (vgl. 1.7). Wir haben hier aber durch Ergebnisse aus dem Deutschen auf Auswirkung bei britisch-englischen Sprechern geschlossen.

Pape und Mooshammer (2006b, o.S.; in 4. ‚Conclusion') schlagen vor, dass es hilfreich wäre, bei zukünftigen Experimenten (sie beziehen sich auf sprachspezifische Unterschiede im Ausmaß der intrinsischen Grundfrequenz und/oder des intrinsic pitchs) die Versuchspersonen, die man untersucht, oder, wie in unserem Fall, Sprecher/Hörer der gleichen Sprache, gleichzeitig auf ihre Sensitivität gegenüber dem *F1-f0*[Bark]-Cue zu testen: „An interesting experiment would be to replicate the F0-F1 distance studies by Traunmüller with listeners of a Romance language to see if F0 actually does not have an influence and therefore does not contribute to the openness perception of vowels, which could be assumed given the presented intrinsic pitch results for Catalan and Italian".

Wie stimmen diesem Anliegen ausdrücklich zu. Streng genommen hätten wir für unsere Perturbationsexperimente Sprecher des Britischen Englisch als Versuchspersonen verwenden sollen, und als Hörer für Perzeptionsexperimente am besten die selben Versuchspersonen, um zu testen, ob wirklich jene, die den *F1-f0*-Cue nutzen, für Vokalhöhenperturbationen auch aktiv mit *f0* kompensieren, oder durch *f0*-Perturbationen ausgelöste Vokalhöhenänderungen durch kompensatorische *F1*-Produktionsänderungen korrigieren.

Dies haben wir nicht getan, sondern diese Effekte im verwandten Deutschen festgestellt. Nun sind diese Ergebnisse hinreichend konsistent mit den Befunden von Vorgängerstudien (1.7, 4.1), so dass wir nicht annehmen müssen, es mit Zufallsbefunden zu tun zu haben, auch wenn natürlich wünschenswert gewesen wäre, wesentlich mehr Versuchspersonen für das Altersstimmenexperiment oder die Perturbationsexperimente zur Verfügung gehabt zu haben. Dass unter diesen wenigen Sprechern dann einige auch noch von den gefundenen und hier beschriebenen Mustern abweichen, also z. B. nur einige wenige tatsächlich *aktiv* unter *F1*-Perturbation *f0* variieren, muss uns allerdings nicht beunruhigen, denn dies – wir haben es in der Einleitung zu dieser Arbeit (1) bereits zitiert – ist laut Hoole und Honda (2011) eher ein Hinweis auf eine aktive Nutzung merkmalsverstärkender Gesten und auf inter-individuelle Unterschiede in der perzeptuellen Sensitivität und Nutzung bestimmter Cues.

So zeigen auch die Daten zu den Änderungen akustischer Parameter zu zwei unterschiedlichen, 20-35 Jahre auseinanderliegenden Zeitpunkten, die wir gegenüber Reubold et al. (2010) um die Daten von drei Sprechern (auf 8) erweitert haben, auch deutliche interindividuelle Unterschiede - nicht bei allen kovariieren *F1* und *f0* in der Weise, wie wir für die in regelmäßigen Abständen über mehrere Jahrzehnte beobachteten zwei Sprecher zeigen konnten. Es könnte sein, dass wir es hier einfach mit zwei Sprechern zu tun hatten, die für den *F1-f0*-Cue besonders sensitiv sind, und andere den Cue weniger bis gar nicht nutzen. Man wird auch zugeben müssen, dass bereits jetzt – 2011 – Veröffentlichungen existieren, die unserer in Reubold et al. (2010) veröffentlichten Kovariationshypothese widersprechen (Rhodes, 2011).

Es mag also sein, dass unsere allgemeinere These, dass altersbedingte Änderungen zu Perturbationen führen können, für die kompensiert werden muss, erst noch durch weitere Daten untermauert werden müsste. Zunächst jedoch, und wahrscheinlich näherliegender, wäre es jedoch wünschenswert, durch artikulatorische Daten unsere These zu untermauern. Die naheliegendste Geste, um *F1* zu beeinflussen, ist, wie erwähnt, eine Variation des Kieferöffnungsgrades. Relativ leicht realisierbar wären also akustische Perturbationsexperimente, bei denen gleichzeitig der Kieferöffnungsgrad aufgezeichnet wird; auch die vertikale Kehlkopflagenvariation, auf die wir aus

unseren akustischen Daten geschlossen hatten, wäre zu untersuchen; wie bereits vorgeschlagen, wäre es ratsam, bei den selben Versuchspersonen die Sensitivität für den *F1-f0*-Cue zu ermitteln.

Es gibt – neben dem hier untersuchten Altern Erwachsener – andere, schneller stattfindende Perturbationen, die sich teilweise sogar auf einzelne Artikulatoren beschränkt ansehen lassen, die man zur weitergehenden Untersuchung möglichen kompensatorischen Verhaltens betrachten könnte.

So ist beispielsweise bekannt, dass das Rauchen die Grundfrequenz absinken lässt, was, wenn das Rauchen aufgegeben wird, sich relativ schnell wieder zurückbilden soll (Verdonck-De Leeuw & Mahieu, 2004). Man könnte also beispielsweise Raucher, die das Rauchen aufgeben wollen, akustisch und artikulatorisch aufnehmen, und in einem gewissen Abstand nach der Aufgabe des Rauchens wieder aufnehmen, um zu testen, ob sich neben der Grundfrequenz auch der erste Formant und der Kieferöffnungsgrad verändert haben.

Wie oben erwähnt, ist eine der ausgeprägtesten, alle Artikulatoren betreffende „Perturbationen“ die Mutation, der Wandel von der Kinder- zur Erwachsenenstimme. Dementsprechend sollte es ein vielversprechender Ansatz sein, akustische wie artikulatorische Daten von Heranwachsenden während der Mutation in regelmäßigen Abständen aufzuzeichnen und auf kompensatorisches Verhalten hin zu überprüfen.

Anhang A

Anhang zu Kapitel 3

A.1 Zusätzliche Abbildungen zu Kapitel 3.2

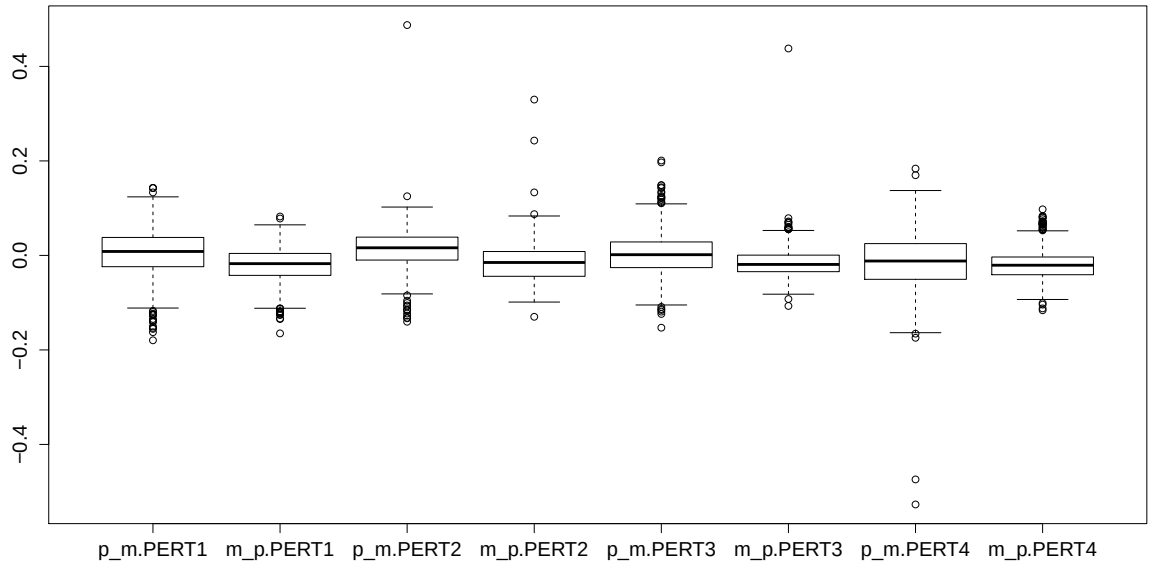

Abbildung A.1: *F1-Perturbation: Unterschiede der Adaptive Response-Werte von f0 für den Faktor SPRECHERGRUPPE, wobei* p_m *die* PLUS-MINUS-*,* m_p *die* MINUS-PLUS-*Gruppe bezeichnet und* PERT1 *bis* PERT4 *die* PERTURBATIONSPHASE *1 bis 4, siehe Fußnote auf Seite 111.*

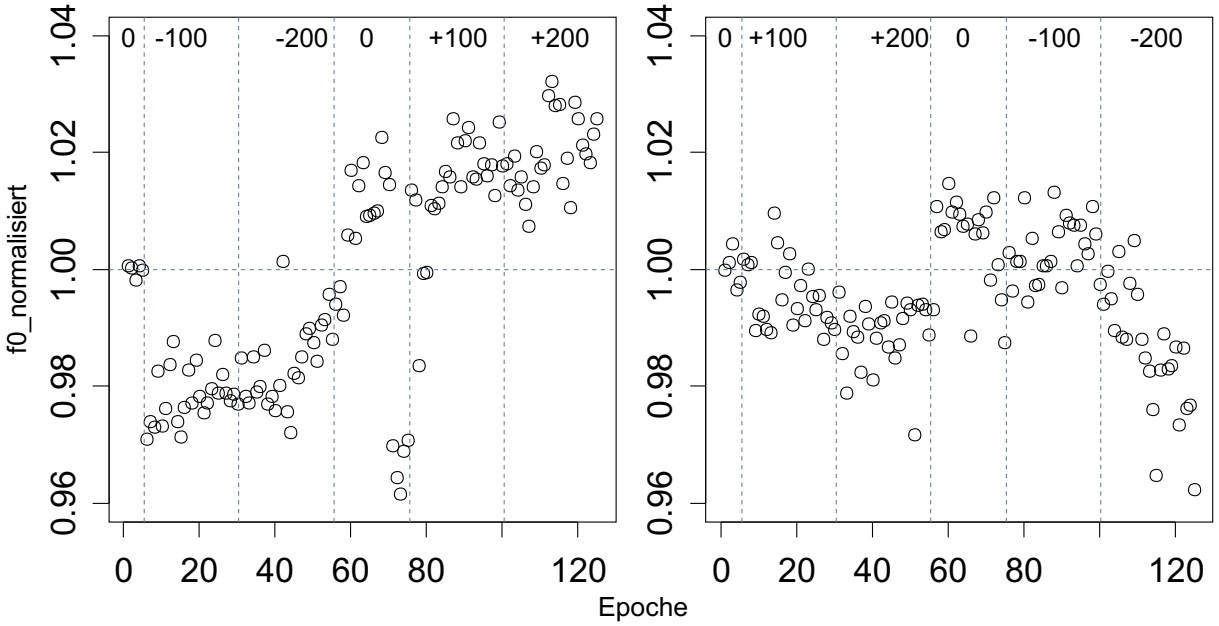

Abbildung A.2: *F1-Perturbation: Über die Sprecher gemittelte, normalisierte f0-Werte (ein Wert pro Epoche), aufgeteilt nach Sprechergruppe (links: MINUS-PLUS, rechts: PLUS-MINUS); die vertikalen Linien trennen die Perturbationsphasen voneinander; siehe Fußnote auf Seite 111.*

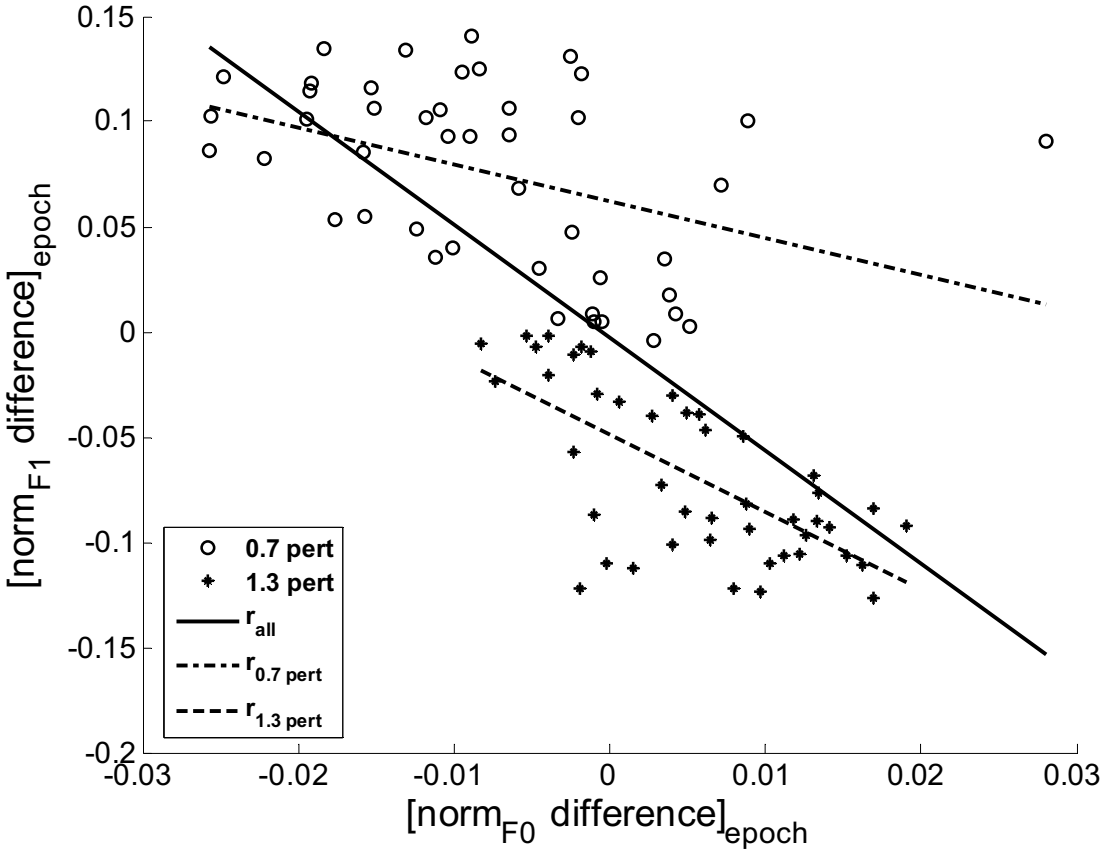

Figure 3.21: Correlation between normalized F0 difference and the normalized F1 difference, over the full pert epochs. The abscissa corresponds to the difference between a subject's normalized F0 and the mean of all subjects' normalized F0, both calculated for a given epoch. The ordinate is the normalized F1 difference, calculated the same way as the F0 difference. The open circles correspond to data from 0.7 pert subjects; the asterisks correspond to data from 1.3 pert subjects. The lines indicate the best regression fit, all of which have significant correlation ($p < 0.001$). The dashed-dotted line corresponds to the correlation of the 0.7 pert data, the dashed line corresponds to the 1.3 pert data, and the solid line corresponds to all data. Only epochs 21 to 65 (full-pert and post-pert epochs) from nineteen subjects are shown; one subject was excluded from this analysis (see text).

Abbildung A.3: *Abbildung aus Villacorta (2006, Seite 66) mit originaler Abbildungsunterschrift. Die Datenpunkte (aufgeteilt nach den zwei Sprechergruppen Villacortas) repräsentieren jeweils eine Epoche, d.h. ein Punkt zeigt den Mittelwert aller Sprecher der entsprechenden Gruppe pro Epoche.*

A.2 Statistikanhang zu Kapitel 3.2

Mit den normierten *F1*- bzw. *f0*-Daten (siehe Formel 3.1) als abhängigen Variablen wurden Lineare Gemischte Modelle berechnet, die, unter Ausklammerung der durch die Sprecher verursachten Variation, den Einfluss der *Experimentphasen* auf die abhängigen Variablen bewertete. Da diese, je nach Sprechergruppe (*MINUSPLUS* oder *PLUSMINUS*), bei erfolgreichen Kompensation für die *F1*-Perturbation in den einzelnen Perturbationsphasen je nach Gruppe mal steigen und mal fallen, und sich daher in beiden Gruppen auftretende Effekte möglicherweise aufheben würden, musste in diesem Fall das Material aufgeteilt werden, und zwar in die Daten der *PLUSMINUS*- und jenen der *MINUSPLUS*-Sprecher (jeweils 7). Die Experimentphasen wurden diesmal wie gegeben angewendet, also ohne Vereinigung der *BASELINE*- mit der *RÜCK*-Phase zu einer *BASIS*-Phase; es gab also für die unabhängige Variable *Experimentphase* sechs Stufen: *BASELINE, PERTURBATION1, PERTURBATION2, RÜCK, PERTURBATION3, PERTURBATION4*. Nachfolgend auf diese Linearen Gemischten Modelle wurden post-hoc-*Tukey*-Tests durchgeführt, um die Paarungen aus den Stufen von *Experimentphase* zu testen.

Für die normalisierten *F1*-Werte in der *MINUSPLUS*-Sprechergruppe ergab sich mit $\chi^2[5] = 305,36; p < 0,001$ ein signifikanter Effekt für *Experimentphase*. Nur die Paarungen der letzten zwei Perturbationphasen zur *BASELINE* waren signifikant (z-Wert für *PERTURBATION3*: $-7,312$[1]$; p < 0,001$, *PERTURBATION4*: $-8,238; p < 0,001$), während die Paarung der *PERTURBATION1* ($z = 0,728$) sowie der *PERTURBATION2*($z = 1,317$) nicht signifikant unterschiedlich von der *BASELINE* waren. Die Paarungen der *RÜCK*-Phase zu den vier Perturbationsphasen ($z : -5,967; -6,929; 7.162$ bzw. $8,674$; mit jeweils $p < 0,001$) unterschieden sich alle voneinander signifikant. *RÜCK*- und *BASELINE*-Phase unterscheiden sich auch signifikant ($z = -2,867; p < 0,05$), aber es gibt keine Unterschiede in den Vergleichen der ersten beiden und letzten beiden Perturbationsphasen (*PERTURBATION2-PERTURBATION1*: $z = 1,020; n.s.$, *PERTURBATION4-PERTURBATION3*: $z = -1,6; n.s.$).

Für die normalisierten *F1*-Werte in der *PLUSMINUS*-Sprechergruppe ergab sich mit $\chi^2[5] = 374,84; p < 0,001$ ein signifikanter Effekt für *Experimentphase*. Nur die Paarungen der letzten zwei Perturbationphasen zur *BASELINE* waren signifikant (z-Wert für *PERTURBATION3*: $6,941; p < 0,001$, *PERTURBATION4*: $8,238; p < 0,001$), während die Paarung der *PERTURBATION1* ($z = -0,234$) sowie der *PERTURBATION2*($z = -1,743$) nicht signifikant unterschiedlich von der *BASELINE* waren. Für die Paarungen der *RÜCK*-Phase zu den vier Perturbationsphasen ($z : 5,966; 8,430; -5,749$ bzw. $-7,869$,; mit jeweils $p < 0,001$) gilt jedoch wieder, dass alle signifikant waren. Für diese Sprechergruppe unterschieden sich auch *RÜCK*- und *BASELINE*-Phase ($z = 3,351; p < 0,01$), aber weder die Paarungen *PERTURBATION4-PERTURBATION3* ($z = 2,248; n.s.$) noch *PERTURBATION2-PERTURBATION1* ($z = -2,614; n.s.$) unterschieden sich signifikant.

Für die normalisierten *f0*-Werte in der *MINUSPLUS*-Sprechergruppe ergab sich mit $\chi^2[5] = 305,36; p < 0,001$ ein signifikanter Effekt für *Experimentphase*. Alle vier Paarungen von Perturbationphasen zur *BASELINE* waren signifikant (z-Wert für *PERTURBATION1*: $-4,627$; *PERTURBATION2*: $-3,457$; *PERTURBATION3*: $3,061$; *PERTURBATION4*: $4,248$; $p < 0,001$ in allen vier Paarungen). Das Gleiche gilt hier auch für die Paarungen der *RÜCK*-Phase zu den vier Perturbationsphasen ($z : 7,069; 5,159; -5,485$ bzw. $-7,424$; mit jeweils $p < 0,001$). *RÜCK*- und

[1] Wir geben die z-Werte hier entgegen sonstiger Gepflogenheiten *mit* Vorzeichen an, um die Richtung der Verschiebung der abhängigen Variable mitzuteilen, ohne Steigungswerte bekanntgeben zu müssen.

BASELINE-Phase unterscheiden sich nicht ($z = -0,292; n.s.$), ebensowenig wie es keine Unterschiede in den Vergleichen der ersten beiden und letzten beiden Perturbationsphasen gibt (*PERTURBATION2-PERTURBATION1*: $z = 2,025; n.s.$, *PERTURBATION4-PERTURBATION3*: $z = 2,057; n.s.$).

Für die normalisierten *f0*-Werte in der *PLUSMINUS*-Sprechergruppe ergab sich mit $\chi^2[5] = 55,65; p < 0,001$ ein signifikanter Effekt für *Experimentphase*. Alle vier Paarungen von Perturbationphasen zur *BASELINE* waren *nicht* signifikant (z-Wert für *PERTURBATION1*: $-0,872$; *PERTURBATION2*: $-2,108$; *PERTURBATION3*: $0,635$; *PERTURBATION4*: $-2,64$; $n.s.$ in allen vier Paarungen). Das Gleiche gilt hier auch für die Paarung der *RÜCK*-Phase zur Perturbationsphase 3 ($0,393; n.s.$), während die anderen Perturbationphasen sich im Vergleich zur *RÜCK*-Phase sehr wohl unterscheiden (*PERTURBATION1*: $z = 2,854; p < 0,05$, *PERTURBATION2*: $z = 4,872; p < 0,001$, *PERTURBATION4*: $z = 5,742; p < 0,001$), wobei, wie man an den Vorzeichen sieht, *f0* beim Ausschalten der Perturbation nach oben nicht sinkt, wie die Hypothese voraussagen würde, sondern stattdessen steigt. *RÜCK*- und *BASELINE*-Phase unterscheiden sich nicht ($z = 0,858; n.s.$), ebensowenig wie es keine signifikanten Unterschiede in dem Vergleich der ersten beiden Perturbationsphasen gibt (*PERTURBATION2-PERTURBATION1*: $z = -2,141; n.s.$), sehr wohl aber im Vergleich *PERTURBATION4-PERTURBATION3*: $z = -5,674; p < 0,001$).

Um diese etwas verwirrende Vielfalt an Ergebnissen zusammenzufassen: Für beide Sprechergruppen gilt, dass *F1* in der ersten und zweiten Perturbationsphase sich nicht von *BASELINE* unterschied, dann aber ein signifikanter Unterschied zur *RÜCK*-Phase einsetzte; die Perturbationsphasen 3 und 4 unterschieden sich dann sowohl von der *RÜCK*-, als auch von der *BASELINE*-Phase, unterschieden sich aber nicht voneinander; die Perturbationsstärke, die ja vom Übergang von der ersten und dritten auf die zweite und vierte Perturbationsphase der entscheidende Unterschied ist, scheint also für *F1* keine Rolle gespielt zu haben. Auch ist bei beiden Gruppen die *BASELINE*- von der *RÜCK*-Phase verschieden, so dass paradoxerweise gesagt werden kann, dass „Kompensation“ im *F1*-Bereich erst dort einsetzt, wo die Perturbation wieder zwischenzeitlich ausgeschaltet wird.

Bezüglich *f0* ist das vereinfachte Bild das folgende: Die *MINUSPLUS*-Sprechergruppe verschiebt, wie in der Hypothese vorausgesagt, die Grundfrequenz immer in Richtung der *F1*-Perturbation. Auch hier machen sie keine Unterschiede, die mit der Perturbationsstärke erklärbar wären. Die *PLUSMINUS*-Gruppe verschiebt *f0* in den ersten beiden Epochen nicht; dann, wenn die *F1*-Perturbation nach oben ausgeschaltet wird, steigt die Grundfrequenz, was nicht in Einklang zu bringen ist mit der Hypothese. In der letzten Perturbationsphase, in der *F1* also nach unten perturbiert wird, erfüllen aber die Sprecher dieser Gruppe wieder die Voraussagen und senken die Grundfrequenz.

Für die Kompensation in *F1* kann also gesagt werden, dass sie in beiden Sprechergruppen zumindest im Mittel über die jeweils sieben Sprecher immer in Gegenrichtung zur *F1*-Perturbation geschieht, zumindest wenn man für die ersten beiden Perturbationsphasen den erstaunlichen Befund akzeptiert, dass es dort keine Änderungen bezüglich zu *BASELINE* gibt, aber sich diese Phasen relativ zur *RÜCK*-Phase so verhalten, wie vorausgesagt. Für die Grundfrequenzbewegung gilt, dass sie für die *MINUSPLUS*-Gruppe tatsächlich immer in Richtung der *F1*-Perturbation verläuft, so wie vorausgesagt, und zwar auch schon bei den ersten beiden Perturbationsphasen, bei denen in *F1* noch nicht kompensiert wird - dies könnte man dahingehend deuten, dass *statt* des ersten Formanten die Grundfrequenz zur Kompensation eingesetzt wird. Dem steht entgegen, dass dies für die *PLUSMINUS*-Gruppe nicht gilt, da diese Sprecher bezüglich der *BASELINE*

gar keine signifikante Änderung aufweisen, aber dafür auf das Ausschalten der Perturbation mit einer starken Reaktion in *f0* antworten. Erst gegen Ende des Experiments, in der letzten Perturbationsphase, verhalten sie sich wie vorausgesagt und verschieben - relativ zur *RÜCK*-Phase - *f0* in Richtung der Perturbation.

Die *MINUSPLUS*-Gruppe antwortet konsistenter, da sie auf unterschiedliche Richtungen der Perturbation konsistent sowohl mit einer Gegenbewegung in *F1* und einer Bewegung in Perturbationsrichtung in *f0* reagiert. Die *PLUSMINUS*-Gruppe ist zumindest bezüglich der Grundfrequenz inkonsistenter, da sie auf beide Perturbationsrichtungen mit einem Abfall der Grundfrequenz reagiert. Daraus ableiten zu wollen, dass es einen echten Effekt der *Präsentationsrichtung* (*MINUSPLUS* vs. *PLUSMINUS*) gäbe, erscheint allerdings bei einer Sprecheranzahl von jeweils nur sieben Personen als nicht valide - individuelle Kompensationsstrategien können hier zu sehr das Bild verzerren.

A.3 Automatische Klassifikation

In diesem Unterkapitel wird automatische Klassifikation angewendet, um die Vokal-Tokens aus dem *F1*-Perturbationsexperiment aus Kapitel 3.2 in zwei Kategorien, nämlich in die unperturbierte und die perturbierte Phase, einteilen zu können. Die Klassifikation wird verschiedene Parameter bzw. Parameterkombinationen testen: erstens das „klassische“ akustische Vokalhöhenkorrelat *F1*, zweitens die Grundfrequenz *f0*, für die gezeigt wurde, dass sie sich unter *F1*-Perturbation in Richtung der Perturbation verschiebt (siehe Kapitel 3.2 sowie Villacorta (2006); Villacorta et al. (2007)) und das sie in der überwiegenden Mehrzahl der Sprachen der Welt mit Vokalhöhe positiv korreliert (siehe u. a. Whalen und Levitt (1995)); drittens wird getestet werden, ob die Klassifikation sich wesentlich verbessert, wenn sowohl *F1* als auch *f0* als Parameter für den automatischen Klassifikator dienen; diese Vermutung beruht auf Vorarbeiten, die eine (je nach Quelle unterschiedlich realisierte) Kombination des ersten Formanten und der Grundfrequenz als im Vergleich zu *F1* besseres Vokalhöhenkorrelat postulieren (Traunmüller, 1981, 1984; Syrdal & Gopal, 1986; J. D. Miller, 1989).

A.3.1 Klassifikation als Mittel zur Bewertung der Relevanz von Parametern

Wie Harrington (2010, Kapitel 9, Seiten 327-380), dessen Methode hier benutzt wird, beschreibt, kann Klassifikation dazu benutzt werden, zu bestimmen, wie effektiv Phonemkategorien durch bestimmte Parameter voneinander unterschieden werden können; hierzu wird ein gegebenes Phon bzw. seine Eigenschaften (z. B. akustische Parameter wie Formantwerte oder spektrale Momente) einer bestimmten Phonemkategorie automatisch zugeordnet, und im Nachhinein kann bestimmt werden, ob diese Klassifikation korrekt war. Auf diese Weise kann man Vergleiche anstellen zwischen den Beiträgen zweier unterschiedlicher Parameter, oder auch, ob die Kombination bestimmter Parameter einzelnen Parametern überlegen ist. Genau diese Anwendung wird in diesem Kapitel von Nutzen sein, um die Frage zu beantworten, wie die Grundfrequenz, der erste Formant, und die Kombination beider Parameter zur Vokalkategorisierung beitragen.

A.3.2 Methodik der automatischen Klassifikation

Um eine solche Klassifikation durchführen zu könne, müssen zunächst Modelle anhand von ausgewählten Parametern von Phonen, deren Kategorienzugehörigkeit bekannt ist, trainiert werden. Anschließend kann diese Modellierung auf unbekannten Phonen zugehörigen Parametern angewendet werden, so dass die Phone aufgrund der Verteilung des gewählten Parameters (oder der gewählten Parameterkombination) bestimmten Phonemkategorien zugeordnet werden. Sofern man die Information, welchen Kategorien diese getesteten Phoneme tatsächlich angehörten, besitzt, kann leicht berechnet werden, wie erfolgreich klassifiziert wurde; in diesem Fall ist das Ergebnis ein Wert zwischen 0 und 1, der die Rate korrekter Klassifikationen angibt.

Diese Verfahren wurde, wie in Harrington (2010) ausführlich beschrieben, in *R* mit Funktionen aus dem package *MASS* (Venables & Ripley, 2002) ausgeführt. Hierzu wird Quadratische Diskriminanzanalyse (Srivastava, Gupta & Frigyik, 2007) angewendet, um auf die Trainingsdaten zu modellieren: für jede bekannte Kategorie wird eine Auftretenswahrscheinlichkeit als Normalverteilung berechnet und aufgrund dieser werden unbekannte Tokens (die Testdaten) einer dieser Kategorien zugeordnet.

Das Verfahren ist, wie alle vergleichbaren Verfahren, immer in Gefahr, übertrainiert zu werden. Sogenanntes „over-fitting“ kann z. B. auftreten, wenn Test- und Trainingsdaten identisch sind, und führt zu Erkennungsraten nahe 1 (oder anders ausgedrückt, nahe 100%), allerdings nur in dem beschriebenen Fall; wirklich unbekanntes Testdatenmaterial kann gegebenenfalls trotz der hohen Erkennungsrate im eben beschriebenen Fall sehr schlecht klassifiziert werden. Daher sollte dieses „over-fitting“ soweit als möglich vermieden werden. Im vorliegenden Fall von Sprachdaten bedeutet dies, dass gewöhnlich vermieden wird, die Daten eines Sprechers sowohl als Trainings-, als auch als Testdaten zu verwenden, zumal, wenn hierzu die gleichen Tokens benutzt werden.

A.3.3 Wiederholung der Klassifikation unter verschiedenen Bedingungen

Wir wollen, um Ergebnisse, die nur zufällig unserer Hypothese entsprechen, zu vermeiden, das Verfahren in mehreren Durchgängen anwenden, wobei wir sowohl die Auswahl der Trainings- und Testdaten als auch die Auswahl der unperturbierten und perturbierten Epochen variieren; hierbei soll jedes Mal vermieden werden, dass Trainings- und Testmaterial identisch sind. In einigen dieser Durchgänge wird die Ähnlichkeit zwischen Trainings- und Testmaterial allerdings durchaus höher sein als in anderen, wodurch in diesen Fällen mit generell höheren Erkennungsraten zu rechnen sein wird. Hauptsache wird jedoch sein, dass das vorhergesagte Muster der Reihenfolge der Parameter/Parameterkombinationen (*f0* als schwächstes , *F1* als starkes, und die Kombination von *F1* und *f0* als stärkstes Korrelat der Kategorienzugehörigkeit), stets bestätigt werden wird. Dies kann auch prüfstatistisch untersucht werden. Bezüglich der Epochenauswahl wollen wir erstens eine gleiche Anzahl perturbierter und unperturbierter Epochen wählen, um die apriori Wahrscheinlichkeit für das Auftreten einer der Kategorien perturbiert/unperturbiert auf je 50% festzulegen; als perturbierte Epochen wählen wir aus den Epochen des Experiments 3.2 jene aus, für die als wahrscheinlich gelten kann, dass die Sprecher überhaupt kompensiert haben, also Epochen aus *PERTURBATIONSPHASE 4*. Wir müssen zu diesem selektiven Verfahren greifen, da es bekannt ist, dass Kompensationen für Perturbationen immer vergleichsweise gering ausfallen, und daher vermutet werden muss, dass zwischen den hier verwendeten Kategorien perturbiert/unperturbiert die Unterschiede geringer ausfallen werden als z. B. zwischen den Vokalen

aus zwei distinkten Phonemkategorien, wie z. B. /iː/ vs. /eː/.

Die einzelnen Testdurchgänge - im folgenden Teilexperiment 1 bis 6 benannt - unterscheiden sich hinsichtlich

... der Auswahl von Trainings- und Testdaten:

Eine der vier Wiederholungen pro Epoche soll als Trainingsdaten verwendet werden und alle anderen Wiederholungen als Testdaten; dieses Verfahren muss für jede der vier Wiederholungen erneut laufen gelassen und die Ergebnisse am Ende gemittelt werden; zwar gibt es keine Daten, die gleichzeitig Trainings- und Testdaten sind, aber die Daten stammen von den selben Sprechern und weisen daher eine Ähnlichkeit auf; daher wird es bei dieser Methode vermutlich zu den höchsten Erkennungsraten kommen.

Die Daten der *PLUS-MINUS*-Sprechergruppe werden als Trainingscorpus verwendet und jene der *MINUS-PLUS*-Sprechergruppe als Testdaten; anschließend wird in umgekehrter Reihenfolge wiederholt und Ergebnisse gemittelt.

Die Daten eines Sprechers werden als Trainings-, die der anderen Sprecher als Testdaten verwendet. Für alle Sprecher wird das Verfahren wiederholt und Ergebnisse gemittelt; wegen der Sprechervariabilität wird dieses Verfahren wahrscheinlich die niedrigsten Erkennungsraten aufweisen.

... der Auswahl der unperturbierten und perturbierten Epochen, wobei anzustreben ist, dass es gleich viele perturbierte wie nicht perturbierte Tokens gibt, so dass die a priori-Wahrscheinlichkeit 50% beträgt:

BASIS- vs. *PERTURBATIONSPHASE 4*-Epochen (wobei *BASIS* die Kombination aus *BASELINE*- und *RÜCK*-Phase darstellt); hierbei pro Kategorie (perturbiert vs. nicht perturbiert) 1400 (=4 (Wiederholungen)*25 (Epochen)*14 (Sprecher)) Tokens (Trainings- und Testdaten zusammen); *PERTURBATIONSPHASE 4* wurde gewählt, da in dieser die Unterschiede zur *BASIS* am deutlichsten waren, siehe Kapitel 3.2

BASELINE-Epochen (=Epochen 1 bis 5) vs. die letzten fünf Epochen von *PERTURBATIONSPHASE 4*; hierbei pro Kategorie 280 (=4*5*14) Tokens (Trainings- und Testdaten zusammen; Aufteilung je nach Wahl der Trainingsdaten)

Es gibt also $2 * 3 = 6$ Klassifikationsexperimente, da die drei gewählten Variationsbedingungen der Trainingscorpusauswahl mit den zwei gewählten Variationsbedingungen der Epochenauswahl durchmischt wurden. Da diese Arten von Klassifikationsexperimenten je dreimal angewendet werden, und zwar erstens mit dem Parameter *Adaptive Response* von *f0*, zweitens mit *Adaptive Response* von *F1*, und drittens mit den zwei Parametern *F1* und *f0* (jeweils transformiert zu *Adaptive Response*-Werten), gehen insgesamt die Ergebnisse von 18 Unterexperimenten in die Untersuchung ein.

Es gibt zwei Möglichkeiten, die Erkennungsraten anzugeben: es kann für jede der Kategorien eine Erkennungsrate bestimmt werden (also n Werte bei n Kategorien), wobei die Erkennungsrate für jede Kategorie angibt, wie viele der unbekannten Testdaten dieser Kategorie korrekt zugeordnet wurden; dies geschieht mittels einer Konfusionsmatrix, die die korrekt erkannten und die inkorrekt zugeordneten Tokens aufzählt; diese Matrix ist dergestalt aufgebaut, dass in der Diagonalen die korrekt erkannten Tokens aufgelistet sind; die kategoriensprezifische Erkennungsrate

Teilexperiment	Trainings-/Testdaten	verwendete Epochen
1	iterativ eine VPn vs. Rest	1-5&121-125
2	iterativ eine VPn vs. Rest	1-5&56-76&101-125
3	iterativ Position x (mit x=1:4) vs. Rest	1-5&121-125
4	iterativ Position x (mit x=1:4) vs. Rest	1-5&56-76&101-125
5	MINUS-PLUS vs. PLUS-MINUS u. u.	1-5&121-125
6	MINUS-PLUS vs. PLUS-MINUS u. u.	1-5&56-76&101-125

Tabelle A.1: *Übersicht über Teilexperimente, Trainings- vs. Testdaten, und verwendete Epochen. Zu 1. und 2.: VPn=Versuchsperson, Rest=alle anderen Versuchspersonen; zu 3. und 4.: Position=Position eines* beten-*Tokens in der Epoche, Rest=alle anderen Positionen; zu 5. und 6.: MINUS-PLUS bzw. PLUS-MINUS sind die zwei Sprechergruppen; zu weiteren Details siehe Kapitel 3.2. Epochenauswahl: 1-5 entspricht der* BASELINE*, 56-75 der* RÜCK-*Phase (jeweils unperturbiert), 101-125 der* PERTURBATIONSPHASE 4*. Weitere Details zur Begründung der Auswahl: siehe Text.*

wird nun so berechnet, dass diese Werte in der Diagonalen, die die Anzahl der korrekten Klassifikationen angibt, durch die Summe der anderen Werte in der die Kategorie betreffenden Spalte, die die Anzahl der Misklassifikationen aufführt, dividiert werden[2](Harrington, 2010, Seiten 350 f.).

Die zweite Möglichkeit ist es, über diese kategorienspezifischen Werte zu mitteln, so dass eine Erkennungsrate pro Teilexperiment und pro getestetem Parameter bzw. Parameterkombination entsteht. So ergeben sich z. B. bei dem Teilexperiment, bei dem eine der beiden Sprechergruppen das Trainingsmaterial, die andere das Testmaterial stellt, und bei dem alle Daten aus den *BASIS*- bzw. der *PERTURBATIONSPHASE 4*-Epochen stammen, bei Verwendung der Parameterkombination *F1&f0* 84.6% korrekt klassifizierter unperturbierter Vokale und 54.7% perturbierte Vokale, die korrekt der Klasse der perturbierten Vokale zugeordnet wurden. Insgesamt jedoch ist der Klassifikationserfolg, oder besser Erkennungsrate, mit 69.7% angegeben, was sich einfach als Mittelwert der oben genannten Werte errechnen lässt, da die a-priori-Wahrscheinlichkeit bei je 50% lag, und daher hier der Mittelwert identisch ist mit der tatsächlichen Rate korrekt zugeordneter Einheiten.

[2]In unserem Fall mit nur zwei Kategorien gibt es natürlich nur einen Wert pro Kategorie für Misklassifikationen (in einer 2×2-Konfusionsmatrix), so dass man sich die erwähnte Summierung ersparen kann

Ergebnisse

In diesem Unterkapitel werden nur die letztgenannten „overall“-Werte dargestellt, da m.E. kein weiterer Erkenntnisgewinn durch die Verwendung kategoriespezifischer Erkennungsraten zu erwarten ist. Abbildung A.4 präsentiert die Verteilung dieser Erkennungsraten, also insgesamt 18 Werte, d.h. je 3 Ergebnisse aus den 6 Teilexperimenten.

Wie vorhergesagt, wirkten sich die Auswahl der Epochen, die in die Untersuchung einflossen, und die Auswahl der Trainingsdaten auf die Erkennungsraten aus. Die „besten“ Erkennungsraten ergaben sich im Teilexperiment, in dem nur die *BASELINE* und die letzten fünf Epochen aus der *PERTURBATIONSPHASE 4* einflossen und eine der vier Positionen in einer Epoche das Trainingsmaterial für die anderen drei Wiederholungen war. Hierbei wurden für *Adaptive Response* von *f0* 70.9% korrekt klassifierte Tokens erreicht, bei Verwendung der *Adaptive Response* von *F1* 73.9%, und bei der kombinierten Verwendung beider Parameter 84.6%. Das Teilexperiment, das am „schlechtesten“ abschnitt, war jenes mit *BASIS* und *PERTURBATIONSPHASE 4* als Daten, und den einzelnen Sprechern als Trainings-, den anderen Sprechern als Testdaten. In diesem Fall wurde, bei Benutzung des Parameters *Adaptive Response* von *f0*, eine Erkennungsrate von 53.2% erreicht, bei Verwendung des ersten Formanten 57.5% und bei der Kombination *F1&f0* 64.8%.

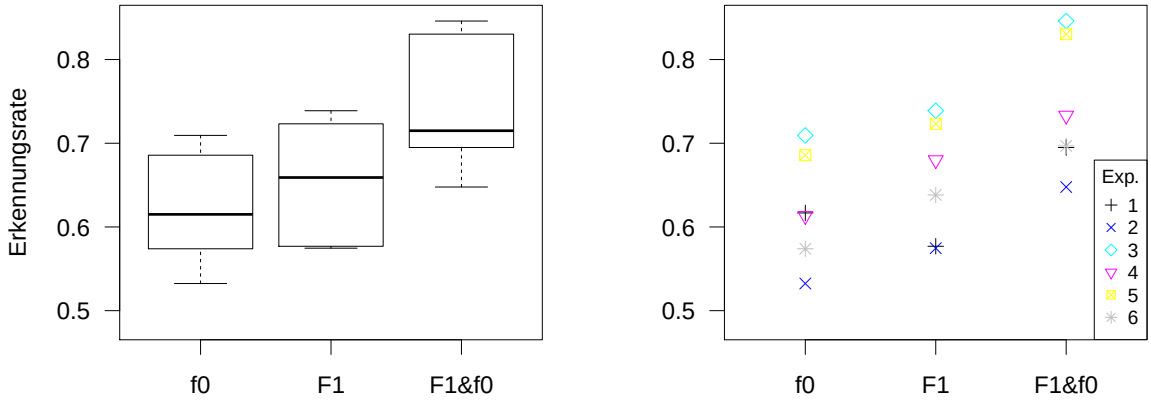

Abbildung A.4: *Links: Verteilung der Erkennungsraten pro Parameter/Parameterkombination. Jeweils sechs Werte pro Boxplot. Rechts: Die gleichen Daten, abgebildet pro Teilexperiment.*

Jede der in Abbildung A.4 gezeigten Verteilungen (Boxen) repräsentiert 6 Werte, nämlich die Ergebnisse der 6 genutzten Möglichkeiten, eine Erkennungsrate pro Parameter zu bestimmen. Man sieht, dass *f0* eine korrekte Klassifikation über Zufallsniveau hinaus ermöglicht, was ein einseitiger Einstichproben-t-Test mit $\mu = 0,5$ bestätigte ($t[5] = 4,49, p < 0,005$); ähnliches gilt in verstärktem Maße auch für *F1* ($t[5] = 5,36, p < 0,005$) und die Kombination *F1&f0* ($t[5] = 7,41, p < 0,001$).

Die Hauptfrage an dieser Stelle ist jedoch, ob sich *f0*, *F1* und die Kombination aus beiden in ihrer Auswirkung auf das Klassifikationsergebnis unterscheiden. Hierbei interessiert es nicht, wie sich die sechs Teilexperimente untereinander unterscheiden, weshalb diese Information ausgeklammert werden muss. Diese Information wurde daher in den Faktor *Teilexperiment* überführt, der in eine ANOVA mit Messwiederholung, die in *R* mit den Funktionen des Packetes *ez* (Lawrence, 2011) berechnet wurde, als error-term einging. Die abhängige Variable war *Erkennungsrate*, der Inner-„Subjekt"-Faktor *Parameter* (mit den Stufen *f0*, *F1* und *F1&f0*; mit „Subjekt" sind hier die Faktorenstufen von *Teilexperiment* (1 bis 6) gemeint). Hierbei konnte ein signifikanter Effekt der unabhängigen Variable *Parameter* ermittelt werden (F[2,10]=48,6; p<0,001). Post-hoc durchgeführte Bonferroni-korrigierte t-Tests für jede der drei möglichen Paarungen - errechnet mit der Funktion *phoc* (Harrington, 2011) - ergab signifikante Unterschiede zwischen *f0* und der *F1&f0*-Kombination ($t[5] = 12,7, p < 0,001$) sowie zwischen *F1* und der *F1&f0*-Kombination ($t[5] = 7,5, p < 0,005$); keine Signifikanz ergab sich für den Unterschied zwischen *f0* und *F1* ($t[5] = 2,1, n.s.$).

Wir konnten also an dieser Stelle zeigen, dass *F1*, aber eben auch *f0* herangezogen werden konnte, um perturbierte von nicht perturbierten Vokalen zu unterscheiden. Allerdings wird man zugeben müssen, dass relativ willkürlich nur die Perturbationsphase 4 am Ende des Experiments herangezogen wurde. Wie Abbildung 3.11 auf Seite 114 zeigt, wurde aber auch schon vorher deutlich für Perturbation des ersten Formanten kompensiert, so z. B. in Perturbationsphase 3. Dort wird relativ stark *F1* von den Sprechern verändert, während die Grundfrequenz vergleichsweise wenig beeinflusst zu sein scheint. Wollte man also das oben beschriebene Experiment wiederholen, indem man Perturbationsphase 4 durch Perturbationsphase 3 ersetzt, ergibt sich möglicherweise kein Vorteil durch Hinzunahme der Grundfrequenz als Klassifikationsparameter.

Aus diesem Grund wurde die oben beschriebene Methode drei weitere Male angewendet, und zwar durch Ersetzung der Perturbationsphase 4 durch die Perturbationsphasen 1 bis 3, wobei bei den Kategorisierungen auf Grundlage der fünf *BASELINE*-Epochen als Vergleichsmaterial stets die jeweils letzten fünf Pherturbationsphasenepochen (also 26 bis 30, 51:55, sowie 96:100) gewählt wurden, so wie auch in der oben genannten Methode. Somit ergeben sich insgesamt für 4 Perturbationsphasen, 3 Parameter und 6 Teilexperimente Ergebnisse für $4 * 3 * 6 = 72$ Werte. Die Ergebnisse werden präsentiert in Abbildung A.5, die den nur mäßigen Klassifikationserfolg aller drei Parameter/Parameterkombination in den ersten 2 Perturbationsphasen deutlich macht. Eine detailierte Auflistung der statistischen Ergebnisse ist im Anhang A.4 zu finden.

An dieser Stelle beschreiben wir hingegen nur die interessantesten Phänomene. Während, wie auch die Statistik in A.4 aufzeigt, in den ersten beiden Perturbationsphasen die Parameterunterschiede zwischen *perturbiert* und *unperturbiert* geringer ausfallen und demensprechend die Klassifikationsergebnisse im Mittel nur gering über Zufallsniveau hinauskommen, sind alle drei Parameter/Parameterkombinationen (*f0*,*F1* und *f0&F1*) so ausgeprägt, dass eine erfolgreiche Klassifikation möglich ist - freilich in unterschiedlichen Ausmaßen. Während *f0* in Perturbationsphase 4 die Klassifikation *perturbiert/unperturbiert* schon bewältigt, ist der Erfolg für *F1* doch erheblich besser, aber nicht so gut wie die Kombination beider Parameter. In der dritten Perturbationsphase hingegen ist der Beitrag der Grundfrequenz geringer, und daher die Kombination beider Parameter auch nicht wesentlich besser als *F1* alleine. Diese wird durch die post-hoc nach einer Varianzanalyse mit Messwiederholung durchgeführten Bonferroni-korrigerten t-Tests, deren Ergebnisse en detail im Anhang in Tabelle A.3 zu finden sind, bestätigt: die Klassifikationserfolge durch Verwendung von *F1* in den Perturbationsphasen 3 und 4 unterscheiden sich nicht

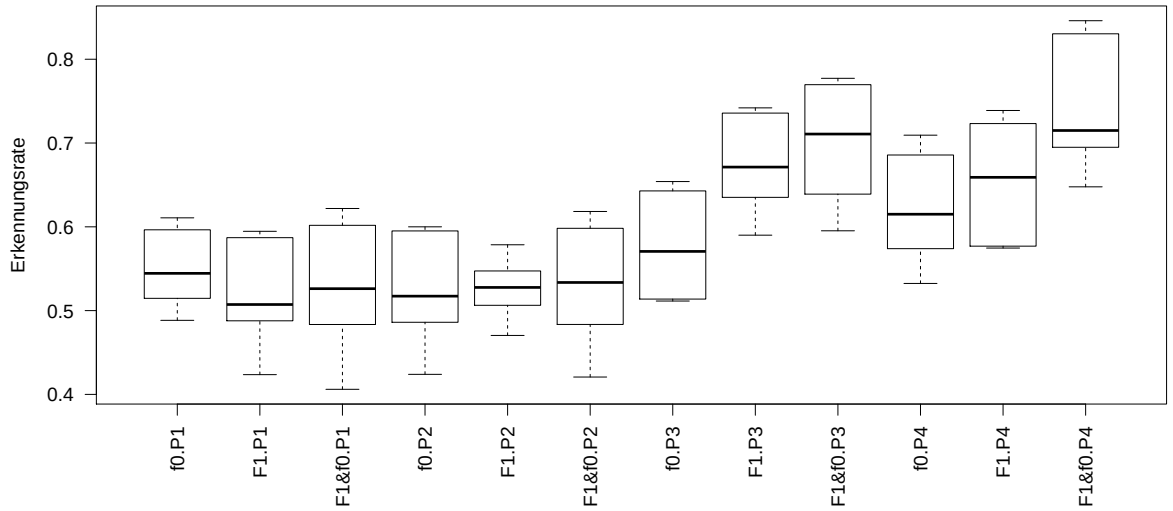

Abbildung A.5: *Verteilung der Erkennungsraten pro Parameter/Parameterkombination und Perturbationsphasen. Pro Box jeweils sechs Werte, gewonnen aus den in A.1 beschriebenen Kategorisierungsverfahren. Die drei Boxen rechts entsprechen denen in Abbildung A.4.*

($t[5] = 2,33; n.s.$), und ebensowenig *F1* von *f0&F1* in Perturbationsphase 3 ($t[5] = 3,4; n.s.$), wohl aber in Perturbationsphase 4 ($t[5] = 7,55; p < 0,05$). Dies ist auch insofern bedeutend, da der Klassifikationserfolg durch *f0* in den beiden Perturbationsphasen sich nur tendenziell, aber nicht signifikant unterscheidet ($t[5] = 4,78; n.s.$).

A.4 Detailierte Statistikergebnisse zum Klassifikationsexperiment

Die nachfolgende Tabelle präsentiert Ergebnisse von t-Tests, mit denen überprüft wurde, ob die Klassifikationsergebnisse der jeweils 6 Teilexperimente pro Parameter/Parameterkombination und pro Perturbationsstufe signifikant größer als 0,5, der dem Zufallsniveau entsprechende Wert, war. Es zeigt sich, dass in den Perturbationsphasen 1 und 2 die Parameterunterschiede zwischen

	PERTURBATION1	**PERTURBATION2**	**PERTURBATION3**	**PERTURBATION4**
f0	$t[5] = 2,58; p < 0,05$	$t[5] = 0,83; n.s.$	$t[5] = 2,89; p < 0,05$	$t[5] = 4,49; p < 0,005$
F1	$t[5] = 0,68; n.s.$	$t[5] = 1,76; n.s.$	$t[5] = 7,21; p < 0,001$	$t[5] = 5,36; p < 0,005$
f0&F1	$t[5] = 0,86; n.s.$	$t[5] = 1,05; n.s.$	$t[5] = 6,84; p < 0,001$	$t[5] = 7,41; p < 0,001$

Tabelle A.2: *Ergebnisse der t-Tests, mit denen für jeden Experimentdurchgang geprüft wurde, ob die Ergebnisse der Klassifikation das Zufallsniveau von 0,5 übersteigen.*

perturbiert und *unperturbiert* nicht stark genug gewesen waren, um zu einer erfolgreichen Klassifikation beizutragen, mit Ausnahme der Grundfrequenz als Parameter in Perturbationsphase 1. In den Phasen 3 und 4 sind alle Werte über Zufallsniveau.

Eine Varianzanalyse mit Messwiederholung mit den Klassifikationsergebnissen als abhängiger Variable, den zwei Inner-Subjekt-Faktoren *Parameter* und *Perturbationsphase*, und dem Faktor *Teilexperiment*, dessen Einfluss ausgeklammert wurde, ergab signifikante Ergebnisse sowohl für *Parameter* ($F[2, 10] = 25, 9; p < 0, 001$), für *Perturbationsphase* ($F[3, 15] = 72, 5; p < 0, 001$) und für die Interaktion beider Faktoren ($F[6, 30] = 16, 9; p < 0, 001$). Daraufhin wurden post-hoc Bonferroni-korrigierte gepaarte t-Tests durchgeführt. Wegen der Menge der Paarungen (66) werden diese in Tabelle A.3 dargestellt:

Paar	t[5]	p	Paar	t[5]	p
f0:P1-F1:P1	2,17	1	f0:P2-F1:P3	-7,08	0,0573
f0:P1-f0&F1:P1	1,22	1	f0:P2-f0&F1:P3	-6,29	0,0982
f0:P1-f0:P2	1,64	1	f0:P2-f0:P4	-6,65	0,0765
f0:P1-F1:P2	2,04	1	f0:P2-F1:P4	-5,05	0,2598
f0:P1-f0&F1:P2	1,19	1	f0:P2-f0&F1:P4	-11,06	0,007
f0:P1-f0:P3	-2,18	1	F1:P2-f0&F1:P2	-0,3	1
f0:P1-F1:P3	-16,77	0,0001	F1:P2-f0:P3	-3,25	1
f0:P1-f0&F1:P3	-10,13	0,0106	F1:P2-F1:P3	-8,4	0,0259
f0:P1-f0:P4	-8,31	0,0273	F1:P2-f0&F1:P3	-7,02	0,0595
f0:P1-F1:P4	-7,49	0,0442	F1:P2-f0:P4	-5,34	0,2037
f0:P1-f0&F1:P4	-13,61	0,0025	F1:P2-F1:P4	-5,62	0,1627
F1:P1-f0&F1:P1	-1,22	1	F1:P2-f0&F1:P4	-9,29	0,0161
F1:P1-f0:P2	-1,58	1	f0&F1:P2-f0:P3	-3,9	0,7495
F1:P1-F1:P2	-0,64	1	f0&F1:P2-F1:P3	-7,14	0,0553
F1:P1-f0&F1:P2	-1,85	1	f0&F1:P2-f0&F1:P3	-6,3	0,0976
F1:P1-f0:P3	-6,17	0,1073	f0&F1:P2-f0:P4	-6,5	0,0852
F1:P1-F1:P3	-7,92	0,0342	f0&F1:P2-F1:P4	-5,21	0,2264
F1:P1-f0&F1:P3	-6,86	0,0665	f0&F1:P2-f0&F1:P4	-11,64	0,0054
F1:P1-f0:P4	-7,12	0,056	f0:P3-F1:P3	-5,63	0,1621
F1:P1-F1:P4	-5,72	0,1505	f0:P3-f0&F1:P3	-5,24	0,2211
F1:P1-f0&F1:P4	-12,03	0,0046	f0:P3-f0:P4	-4,78	0,3285
f0&F1:P1-f0:P2	0,57	1	f0:P3-F1:P4	-3,32	1
f0&F1:P1-F1:P2	0,07	1	f0:P3-f0&F1:P4	-9,93	0,0117
f0&F1:P1-f0&F1:P2	-1,21	1	F1:P3-f0&F1:P3	-3,41	1
f0&F1:P1-f0:P3	-3,77	0,861	F1:P3-f0:P4	5,33	0,2049
f0&F1:P1-F1:P3	-6,55	0,0822	F1:P3-F1:P4	2,33	1
f0&F1:P1-f0&F1:P3	-5,98	0,1239	F1:P3-f0&F1:P4	-6,34	0,095
f0&F1:P1-f0:P4	-5,89	0,132	f0&F1:P3-f0:P4	5,1	0,248
f0&F1:P1-F1:P4	-4,97	0,2777	f0&F1:P3-F1:P4	6,05	0,1172
f0&F1:P1-f0&F1:P4	-11,13	0,0067	f0&F1:P3-f0&F1:P4	-3,29	1
f0:P2-F1:P2	-0,21	1	f0:P4-F1:P4	-2,11	1
f0:P2-f0&F1:P2	-1,1	1	f0:P4-f0&F1:P4	-12,69	0,0036
f0:P2-f0:P3	-6,49	0,0852	F1:P4-f0&F1:P4	-7,55	0,0427

Tabelle A.3: *Post-hoc Bonferroni-korrigierte gepaarte t-Tests für die 66 möglichen Paarungen von* Parameter *und* Perturbationsphase.

Anhang B

Anhang zu Kapitel 4

B.1 Tabelle Sprachmaterial

Prompt	Zielvokal	Kontext	Perzeptionsexperiment
Ich habe bieten gesagt	iː	/bVt/	ja
Ich habe bitten gesagt	ɪ	/bVt/	ja
Ich habe Bütten gesagt	ʏ	/bVt/	nein
Ich habe beten gesagt	eː	/bVt/	ja
Ich habe betten gesagt	ɛ	/bVt/	ja
Ich habe bäten gesagt	ɛː	/bVt/	nein
Ich habe baten gesagt	aː	/bVt/	nein
Ich habe boten gesagt	oː	/bVt/	nein
Ich habe booten gesagt	uː	/bVt/	nein
Ich habe Riege gesagt	iː	/ʀVg/	nein
Ich habe Rüge gesagt	yː	/ʀVg/	nein
Ich habe rege gesagt	eː	/ʀVg/	nein
Ich habe rage gesagt	aː	/ʀVg/	nein
Ich habe riefe gesagt	iː	/ʀVf/	nein
Ich habe Riffe gesagt	ɪ	/ʀVf/	nein
Ich habe reffe gesagt	ɛ	/ʀVf/	nein
Ich habe raffe gesagt	a	/ʀVf/	nein
Ich habe rufe gesagt	uː	/ʀVf/	nein
Ich habe Ziege gesagt	iː	/tsVg/	nein
Ich habe Züge gesagt	yː	/tsVg/	nein
Ich habe Zuge gesagt	uː	/tsVg/	nein

Tabelle B.1: Das von einem 34-jährigen Sprecher des Standarddeutschen gelesene Sprachmaterial; der Satzakrenzt wurde jeweils auf den Zielwörtern produziert; dargestellt sind Zielvokal und Kontext; die letzte Spalte markiert jene Wörter, die zur Kontinuaerstellung herangezogen wurden.

B.1.1 Perzeptionsergebnisse der Lautsprecherhörer

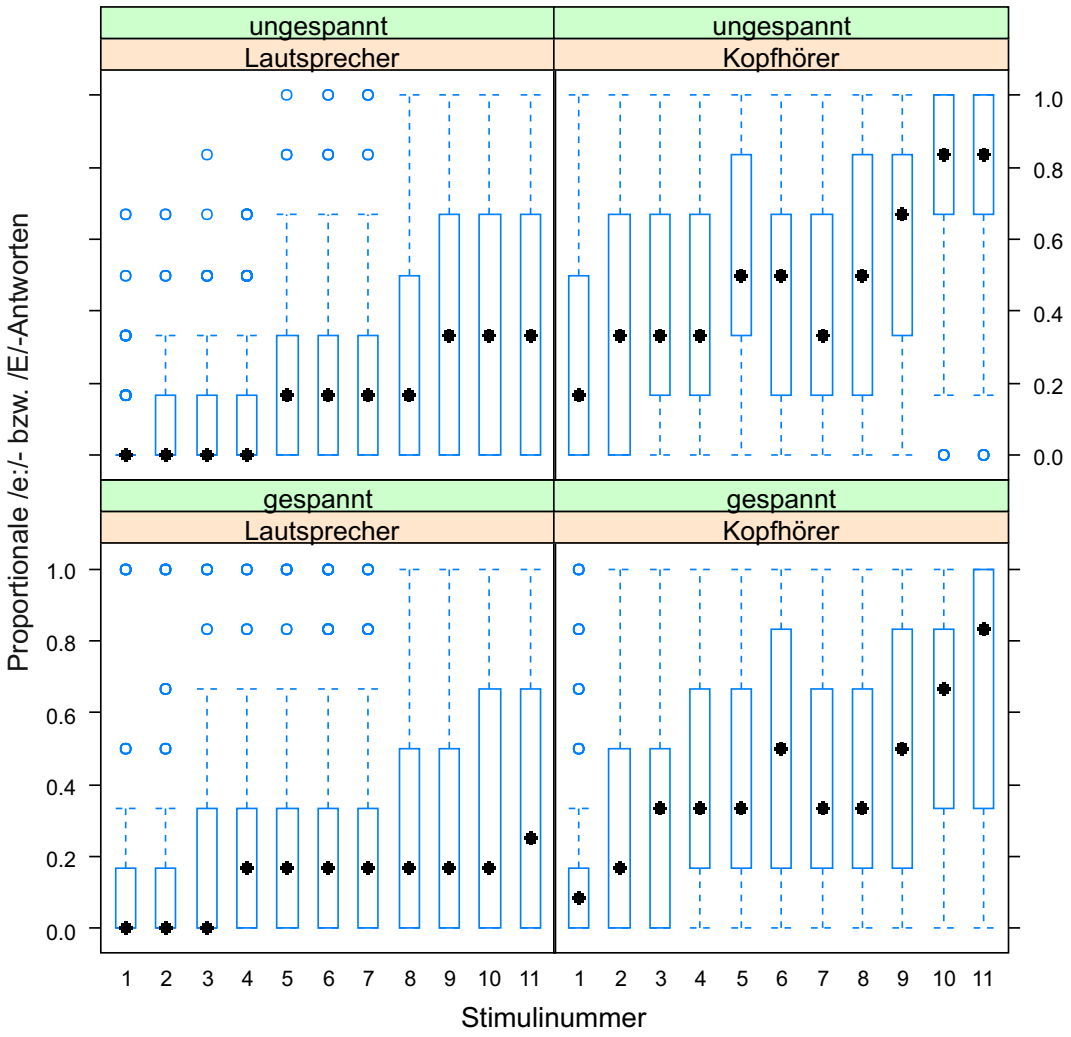

Abbildung B.1: *Einfluss des Audioequipments (*Lautsprecher *links,* Kopfhörer *rechts) auf die Vokalkategorisierung, getrennt nach Gespanntheit (*gespannt *unten,* ungespannt *oben), aber gemittelt über die zwei Kontinuumstypen (*globale *vs.* lokale *f0-Manipulation) dargestellt.*

Abbildung B.1 zeigt die durch die Verwendung von Lautsprechern oder von Kopfhörern verursachten Unterschiede in den Antworten der Versuchspersonen auf die unterschiedlichen Kontinua (hier gemittelt über die Manipulationstypen *global* vs. *lokal*). Wie in Kapitel 4.2.8 festgestellt wurde, hatte der Faktor *Darbietung* (mit den Stufen *Lautsprecher* und *Kopfhörer*) einen signifikanten Einfluss auf die Antworten ($\chi^2[4,5] = 5,88; p < 0,05$, ermittelt über ein Generalisiertes Lineares Gemischtes Modell mit den proportionalen Antworten als abhängiger Variable, *Stimulus* und *Darbietung* als unabhängige Variablen, und unter Ausklammerung der Variation, die durch die unterschiedlichen Hörer bedingt wurde). Dieser festgestellt Unterschied besteht darin, dass die Hörer, die die Stimuli über Lautsprecher dargeboten bekamen, Schwierigkeiten hatten, überhaupt

Unterschiede der Vokalqualität in diesem grundfrequenzverschobenen Kontinuum wahrzunehmen, was Abbildung B.2 zeigt. Zwar wandert auch hier die Identifikationskurve beim Abfall der Grundfrequenz zunehmend in Richtung von /eː/- bzw. /ɛ/-Antworten, erreicht aber im Mittel in keinem der Kontinua einen Umkipppunkt, d.h. immer bleiben die Antworten für jeden Stimulus in der Mehrheit bei /iː/ bzw. /ɪ/.

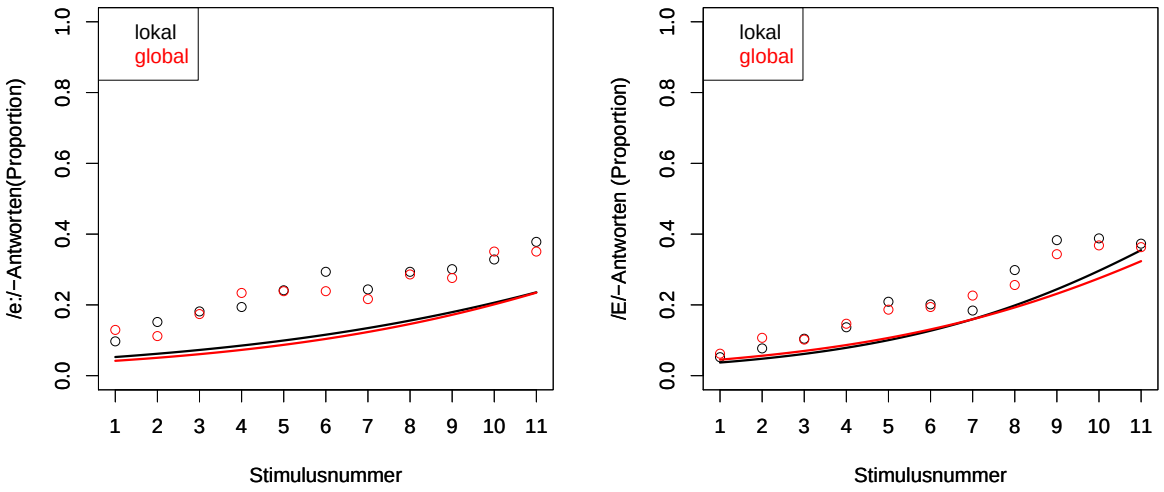

Abbildung B.2: *Identifikationskurven der Lautsprecherhörer zu je zwei Kontinua von* Ich habe *bieten* gesagt *zu* Ich habe *beten* gesagt *(linke Abbildung) bzw. von* Ich habe *bitten* gesagt *zu* Ich habe *Betten* gesagt*, wobei die Grundfrequenz von Stimulus 1 bis Stimulus 11 um fünf Halbtöne* lokal *(schwarz) bzw.* global *(rot) variiert wurde.*

Dennoch lassen sich - rein rechnerisch - die Fragen aus Kapitel 4.1 beantworten. Auch im Falle der Lautsprecherhörer erklären alle vier Kontinua die Variation der proportionalen Antworten, wenn man jeweils ein Generalisiertes Lineares Gemischtes Modell mit der unabhängigen Variable *Stimulusnummer* rechnet, die die abhängige Variable *proportionale Antworten* modelliert, und zwar in dem man den Einfluss der unterschiedlichen Hörer wie üblich ausklammert (bieten-beten-lokal: $\chi^2[1] = 13,9; p < 0,001$; bieten-beten-global: $\chi^2[1] = 18,2; p < 0,00$, bitten-Betten-lokal: $\chi^2[1] = 36,3; p < 0,001$; bitten-Betten-global: $\chi^2[1] = 36,0; p < 0,001$), was an sich, es muss hier wiederholt werden, relativ wenig über den Einfluss der Kontinua aussagt, außer, dass eine gewisse Variation durch die *Stimulusnummer* beeinflusst wird. Über die Kategorisierung durch die Hörer und durch die unterschiedlichen Kontinua verursachten Unterschiede ist hierbei noch nichts gesagt.

Wie die Abbildung B.2 zeigt, gibt es im Mittel keinen Umkipppunkt. Wenn man Umkipppunkte pro Hörer errechnet, wird man feststellen, dass es durchaus Versuchspersonen gab, die - durch die *f0*-Variation in den Stimulusstufen beeinflusst, zwei Kategorien wahrnahmen. Diese Versuchspersonen stellen allerding eine Minderheit. Von den 67 Hörern haben nur 23 bei bieten-beten-lokal, 21 bei bieten-beten-global, 29 bei bitten-Betten-lokal und 29 bei bitten-Betten-global einen Umkipppunkt im Bereich zwischen Stimulus 1 und 11. Ein Vergleich der Einflüsse auf die Stei-

gungen der individuellen Identifikationsstufen mit einer Varianzanalyse mit Messwiederholung, die wieder die Hörer ausklammert, und die Steigungswerte als abhängige Variable und die zwei Zwischen-Subjekt-Faktoren *Gespanntheit* und *Typ der Verschiebung* aufweist, ergibt einen signifikanten Effekt für *Gespanntheit* ($F[1, 66] = 16, 3; p < 0, 001$), aber nicht für *Typ der Verschiebung* ($F[1, 66] = 0, 27; n.s.$). Die Interaktion beider Faktoren ist signifikant ($F[1, 66] = 7, 0; p < 0, 05$); post-hoc-$t$-Tests mit Bonferroni-Korrektur zeigen jedoch lediglich für den *lokalen* Verschiebungstyp Unterschiede, die durch die *Gespanntheit* verursacht wurden ($t[66] = 4, 6; p < 0, 001$), alle anderen Vergleiche ergeben Insignifikanzen.

Dies wird auch bestätigt durch Generalsisierte Lineare Gemischte Modelle für die proportionalen Antworten der Hörer als abhängiger Variable und *Stimulus* sowie *Gespanntheit* als unabhängige Variablen, unter Ausklammerung der hier unerheblichen Variation, die durch die Hörer als auch durch die *Typen der Verschiebung* bedingt ist. Auch hier ergibt sich eine signifikante Beeinflussung der Antworten durch *Gespanntheit* ($\chi^2[1] = 57, 6; p < 0, 001$)[1]. Wie Abbildung B.2 zeigt, ist die Identifikationskurve für die ungespannten Kontinua steiler; dies bedeutet im Umkehrschluss, dass die Antworten in den *gespannt*-Kontinua weniger von *Stimulus* beeinflusst wurden als die der *ungespannt*-Kontinua.

[1] Über die Problematik dieser Art von Vergleich, die dadurch verursacht ist, dass die Kontinua zwar in Schrittweite und Umfang, aber nicht in der Lage identisch sind, sondern ein Verschiebung um circa einen viertel Halbton aufweisen, ist in Kapitel 4.3 bereits berichtet worden.

Anhang C

Publikationsliste gemäß der dritten Satzung zur Änderung der Promotionsordnung der Ludwig-Maximilians-Universität München für die Grade Dr. phil. und Dr. rer. pol. vom 19. Juni 2009

Gemäß der oben genannten dritten Satzung zur Änderung der Promotionsordnung der Ludwig-Maximilians-Universität München für die Grade Dr. phil. und Dr. rer. pol. vom 19. Juni 2009 sind jene Teile der Dissertation zu benennen, die bereits vorab als Beiträge erschienen sind. Desweiteren sind sämtliche Veröffentlichungen des Autors der Dissertation anzugeben, wenn dieser Beitrag nicht vom Autoren alleine verfasst wurde.

Für das erste experimentelle Kapitel 2 trifft zu, dass es aus einem Beitrag entstanden ist, auch wenn Teile des Beitrags hier weggelassen wurden, und dafür neue Teile hinzugefügt wurden. Es handelt sich um folgende Veröffentlichung:

> Reubold, Ulrich, Jonathan Harrington & Felicitas Kleber. 2010. Vocal aging effects on F0 and the first formant: A longitudinal analysis in adult speakers. *Speech Communication, 52*(7-8), 638-651.

Weitere Arbeiten, bei denen der Autor vor Abgabe der Dissertationsschrift lediglich als Mitautor mitgewirkt hat, sind:

Reubold, Ulrich & Alexander Steffen. 2005. Pitch-effects in Diphone recordings: are logatomes inappropriate? In: *Proceedings of INTERSPEECH, Lisbon*, pp. 2797-2800.

Harrington, Jonathan, Felicitas Kleber & Ulrich Reubold. 2007. U-fronting in RP: a link between sound change and diminished perceptual compensation for coarticulation? In: *Proceedings of the International Congress of Phonetic Sciences, Saarbrücken*, pp. 1473-1476.

Harrington, Jonathan, Felicitas Kleber & Ulrich Reubold. 2008. Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study. *The Journal of the Acoustical Society of America, 123*(5), 2825-2835.

Harrington, Jonathan, Felicitas Kleber & Ulrich Reubold. 2008. The acoustic and perceptual bases of diachronic u:-fronting in Standard Southern British. *The Journal of the Acoustical Society of America, 123*(5), 3068.

Kleber, Felicitas, Ulrich Reubold & Jonathan Harrington. 2010. /u/-fronting in RP and the implications of perceptual integration of lip gestures for sound change processes. (Abstract). In: *Abstractbook Laboratory Phonology 12, Albuquerque, New Mexico.*

Müller, Viola, Jonathan Harrington, Felicitas Kleber & Ulrich Reubold. 2011. Age-dependent differences in the neutralization of the intervocalic voicing contrast: Evidence from an apparent-time study on East Franconian. In: *Proceedings of the 12th Annual Conference of the International Speech Communication Association (Interspeech2011)*, pp. 633-636.

Harrington, Jonathan, Phil Hoole, Felicitas Kleber & Ulrich Reubold. 2011. The physiological, acoustic, and perceptual basis of high back vowel fronting: Evidence from German tense and lax vowels. *Journal of Phonetics, 39*(2). 121-131.

Harrington, Jonathan, Felicitas Kleber & Ulrich Reubold. 2011. The contributions of the lips and the tongue to the diachronic fronting of high back vowels in Standard Southern British English. *Journal of the International Phonetic Association, 41*(2), 137-156.

Kleber, Felicitas, Jonathan Harrington & Ulrich Reubold. 2012. The relationship between the perception and production of coarticulation during a sound change in progress. *Language and speech, 55*(3), 383-405.

Danksagung

An erster Stelle möchte ich meinem Doktorvater Jonathan Harrington Dank aussprechen, für die Anregung zu dieser Arbeit, für die hervorragende Betreuung während meines Promotionsstudiums, für Freiräume, die er mit gelassen hat, und die Geduld, die er insbesondere in der letzten Phase aufbringen musste. Ohne ihn wäre auch das gemeinsame Paper, das in großen Teilen hier in das erste experimentelle Kapitel eingegangen ist, nicht so geworden, wie es ist, und wäre wohl auch nicht so schnell veröffentlicht worden.

Auch bei Phil Hoole möchte ich mich herzlich bedanken – nicht nur habe ich sehr viel von ihm während meines Studiums gelernt, sondern auch während der letzten Jahre. Erstaunlich ist, dass ich bis vor nicht allzu langer Zeit von seinen Arbeiten, die die aktive Nutzung/Verstärkung der intrinsische Grundfrequenz betreffen – und somit in gewisser Weise mit der Sensitivität einzelner Hörer auf den *F1-f0*-Cue –, nichts wusste. Diese und verwandte Arbeiten haben diese Dissertation in teilweise neue Bahnen gelenkt, denn meine Ergebnisse passten wunderbar zu seinen (und zu denen der verwandten Arbeiten, denen ich vorher auch nicht soviel Wichtigkeit beimaß). Dennoch hoffe ich, mich nicht allzusehr daran „angelehnt“ zu haben. Phil Hoole ist außerdem u. a. auch zu verdanken, dass das Formantperturbationsexperiment doch recht schnell starten konnte.

Felicitas Kleber danke ich auch, denn sie ist eine Kollegin, Kommilitonin, und Mitstreiterin, wie man sie sich nicht besser wünschen kann. Ihre Tür stand mir bei Diskussionsbedarf immer offen, und diese Diskussionen waren von unschätzbarem Wert.

Da wir schon bei Kommilitonen sind: Allen Teilnehmern des Doktoranden-/Postdoktorandenseminars sei herzlich gedankt! Insbesondere von den alten Hasen kommen immer wieder Nachfragen und Anregungen, die man nicht missen möchte.

Lasse Bombien sei gedankt, dass er stets so hilfsbereit war. Insbesondere hat er auch seine Stimme hergegeben für die in dieser Arbeit vorgestellten Perzeptionsexperimente.

Ohne Christoph Draxler wäre ich nicht so schnell und so einfach an so viele Daten so vieler Hörer in diesen Perzeptionsexperimenten herangekommen. Auch hierfür vielen Dank!

Den irrsinnig komplizierten Aufbau für die Perturbationsexperimente im Studio des IPS hätte ich ohne Klaus Jänsch niemals so hinbekommen. Ohne Klaus Jänsch hätte ich auch manch anderes technisches Problem nicht in den Griff bekommen.

Shanqing Cai danke ich für die Überlassung des Formantperturbationsalgorithmus, sowie der schnellen Lösung damit aufgetretener Probleme.

Meine Familie ist im letzten Jahr der Bearbeitung dieser Dissertation gleich von mehreren Schicksalschlägen getroffen worden. Umso erstaunlicher ist es, dass ich von ihren Mitgliedern gerade auch in dieser Phase soviel Unterstützung und Verständnis entgegen nehmen durfte.

Meiner Frau, Olga Dioubina-Reubold, danke ich ganz besonders für ihre Unterstützung und Geduld, die sie tagtäglich aufbringen musste, für ihre zahlreichen Ratschläge in Bezug auf einige

technische Fragen (so habe ich z. B. von ihr LaTeX und *praat*-Scripting gelernt), und für ihre Bereitschaft, trotz allem ihr Leben mit mir zu teilen. Dies ist für mich von unschätzbarem Wert.

Literaturverzeichnis

Abbs, J. H. & Gracco, V. L. (1984). Control of complex motor gestures: Orofacial muscle responses to load perturbations of lip during speech. *Journal of Neurophysiology*, *51* (4), 705-723.

Abitbol, J., Abitbol, P. & Abitbol, B. (1999). Sex hormones and the female voice. *Journal of Voice*, *13* (3), 424-446. Zugriff auf `http://www.sciencedirect.com/science?_ob=GatewayURL&_origin=ScienceSearch&_method=citationSearch&_piikey=S0892199799800484&_version=1&_returnURL=&md5=917168c7c53b8cf8a71987557b09e55d`

Adank, P. M. (2003). *Vowel normalization: a perceptual-acoustic study of Dutch vowels* (Unveröffentlichte Dissertation). Katholieke Universiteit Njimegen, Njimegen, Niederlande.

Adank, P. M., Smits, R. & van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *The Journal of the Acoustical Society of America*, *116* (5), 3099-3107. Zugriff auf `http://link.aip.org/link/?JAS/116/3099/1`

Ainsworth, W. A. (1975). Intrinsic and extrinsic factors in vowel judgments. In G. Fant & M. Tatham (Hrsg.), *Auditory analysis and perception of speech* (S. 103-113). London, Vereinigtes Königreich: Academic Press.

Assmann, P. F. & Nearey, T. M. (2008). Identification of frequency-shifted vowels. *The Journal of the Acoustical Society of America*, *124* (5), 3203-3212. Zugriff auf `http://link.aip.org/link/?JAS/124/3203/1/http://dx.doi.org/10.1121/1.2980456`

Assmann, P. F., Nearey, T. M., Bharadwaj, S. V., Hubbard, D. & Jayaraman, A. (2008). Developmental study of the relationship between F0 and formant frequencies. *The Journal of the Acoustical Society of America*, *124* (4), 2556. Zugriff auf `http://link.aip.org/link/?JAS/124/2556/1`

Assmann, P. F., Nearey, T. M. & Hogan, J. T. (1982). Vowel identification: orthographic, perceptual, and acoustic aspects. *The Journal of the Acoustical Society of America*, *71* (4), 975-989.

Atkinson, J. E. (1978). Correlation analysis of the physiological factors controlling fundamental voice frequency. *Journal of the Acoustical Society of America*, *63* (1), 211-222.

Auer, P., Barden, B. & Großkopf, B. (1993). Dialektwandel und sprachliche Anpassung bei „Übersiedlern“ und „Übersiedlerinnen“ aus Sachsen. *Deutsche Sprache*, *21*, 80-87.

Auteserre, D., Di Cristo, A. & Hirst, D. J. (1986). Approche physiologique des intonations de base du français. In *Prodeedings of the 15th Journée d'Etude de la Parole (Aix-en-Provence)* (S. 37-41). Aix-en-Provence, Frankreich.

Awan, S. N. & Mueller, P. B. (1992). Speaking fundamental frequency characteristics of centenarian females. *Clinical Linguistics & Phonetics*, *6* (3), 249-254. Zugriff auf http://informahealthcare.com/doi/abs/10.3109/02699209208985533

Baayen, R. H. (2009). *Analyzing linguistic data: A practical introduction to statistics using R* (1. Aufl.). Cambridge [u.a.], Vereinigtes Königreich: Cambridge University Press.

Badin, P. & Fant, G. (1984). Notes on vocal tract computation. In *Speech transmission laboratory, quarterly progress status report* (Bd. 2-3, S. 53-108). Stockholm, Schweden: Department of Speech, Music, and Hearing, KTH, Stockholm.

Baken, R. J. (2005). The Aged Voice: A New Hypothesis. *Journal of Voice*, *19* (3), 317-325. Zugriff auf http://www.sciencedirect.com/science?_ob=GatewayURL&_origin=ScienceSearch&_method=citationSearch&_piikey=S0892199704001055&_version=1&_returnURL=&md5=3c8c89f90500a31b7f457e1f2f8f292d

Barney, A., de Stefano, A. & Henrich, N. (2007). The effect of glottal opening on the acoustic response of the vocal tract. *Acta Acustica united with Acustica, 93, (6), 1046-1056.*.

Bates, D., Maechler, M. & Bolker, B. (2011). *lme4: Linear mixed-effects models using S4 classes.* Zugriff auf http://CRAN.R-project.org/package=lme4

Benjamin, B. J. (1981). Frequency variability in the aged voice. *Journal of gerontology*, *36* (6), 722-729.

Beringer, N. & Schiel, F. (1999). Independent automatic segmentation of speech by pronunciation modeling. In *Proceedings of the International Congress of Phonetic Sciences (ICPhS).* San Francisco, USA.

Berry, J. & Moyle, M. (2011). Covariation among vowel height effects on acoustic measures. *The Journal of the Acoustical Society of America*, *130* (5), EL365-EL371. Zugriff auf http://dx.doi.org/10.1121/1.3651095

Beyerlein, P., Cassidy, A., Kholhatkar, V., Lasarcyk, E., Nöth, E., Potard, B., ... Xu, P. (2008). *Vocal Aging Explained by Vocal Tract Modelling: 2008 JHU Summer Workshop Final Report.* Zugriff auf http://www.clsp.jhu.edu/workshops/ws08/documents/ws08vaeFinalReport.pdf

Biever, D. M. & Bless, D. M. (1989). Vibratory characteristics of the vocal folds in young adult and geriatric women. *Journal of Voice*, *3* (2), 120-131. Zugriff auf http://www.sciencedirect.com/science/article/B7585-4GSCJKH-5/2/b67dc07312c692b00e87c3b1e8208d44

Black, J. W. (1961). Relationships among fundamental frequency, vocal sound pressure, and rate of speaking. *Language & Speech*, *4* (1), 196-199. Zugriff auf http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=15504861&site=ehost-live

Bloothooft, G. & Plomp, R. (1985). Spectral analysis of sung vowels. II. The effect of fundamental frequency on vowel spectra. *The Journal of the Acoustical Society of America*, *77* (4), 1580-1588. Zugriff auf http://dx.doi.org/10.1121/1.392001

Boersma, P. & Weenink, D. (2010). *Praat: doing phonetics by computer [Computer program].* Zugriff am 25.07.2011 auf http://www.fon.hum.uva.nl/praat/

Bradlow, A. R. (2002). Confluent talker-and listener-oriented forces in clear speech production. In C. Gussenhofen & N. Warner (Hrsg.), *Papers in Laboratory Phonology VII* (Bd. 1, S. 241-273). New York, USA: Mouton de Gruyter.

Bresch, E. & Narayanan, S. (2010). Real-time magnetic resonance imaging investigation of resonance tuning in soprano singing. *Journal of the Acoustical Society of America*, *128* (5), EL335-EL341. Zugriff auf http://scitation.aip.org/getpdf/servlet/GetPDFServlet?filetype=pdf&id=JASMAN0001280000050EL335000001&idtype=cvips&doi=10.1121/1.3499700&prog=normal&bypassSSO=1

Brown, W. S., Morris, R. J., Hollien, H. & Howell, E. (1991). Speaking fundamental frequency characteristics as a function of age and professional singing. *Journal of Voice*, *5* (4), 310-315. Zugriff auf http://dx.doi.org/10.1016/S0892-1997(05)80061-X

Brückl, M. (2007). Women's Vocal Aging: a Longitudinal Approach. In *Proceedings of INTERSPEECH 2007, Antwerp* (S. 1170-1173). Antwerpen, Belgien.

Brunner, J., Ghosh, S., Hoole, P., Matthies, M., Tiede, M. & Perkell, J. (2011). The Influence of Auditory Acuity on Acoustic Variability and the Use of Motor Equivalence During Adaptation to a Perturbation. *Journal of Speech, Language, and Hearing Research*, *54* (3), 727-739.

Burnett, T. A., Freedland, M. B., Larson, C. & Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. *The Journal of the Acoustical Society of America*, *103* (6), 3153-3161. Zugriff auf http://dx.doi.org/10.1121/1.423073

Cai, S., Boucek, M., Ghosh, S. S., Guenther, F. H. & Perkell, J. S. (2008). A System for Online Dynamic Perturbation of Formant Trajectories and Results from Perturbations of the Mandarin Triphthong /iau/. In *Proceedings the 8th international seminar on speech production 2008* (S. 65-68). Straßburg, Frankreich: INRIA. Zugriff auf http://issp2008.loria.fr/Proceedings/PDF/issp2008-10.pdf

Cai, S., Ghosh, S. S., Guenther, F. H. & Perkell, J. S. (2010). Adaptive auditory feedback control of the production of formant trajectories in the Mandarin triphthong /iau/ and its pattern of generalization. *The Journal of the Acoustical Society of America*, *128* (4), 2033-2048. Zugriff auf http://link.aip.org/link/?JAS/128/2033/1/http://dx.doi.org/10.1121/1.3479539

Carlsson, G. & Sundberg, J. (1992). Formant frequency tuning in singing. *Journal of Voice*, *6* (3), 256-260. Zugriff auf http://dx.doi.org/10.1016/S0892-1997(05)80150-X

Casserly, E. D. (2011). Speaker compensation for local perturbation of fricative acoustic feedback. *The Journal of the Acoustical Society of America*, *129* (4), 2181-2190. Zugriff auf http://dx.doi.org/doi/10.1121/1.3552883

Cavelaars, A. E. J. M., Kunst, A. E., Geurts, J. J. M., Crialesi, R., L.Grotvedt & Helmert, U. (2000). Persistent variations in average height between countries and between socio-economic groups: an overview of 10 European countries. *Annals of Human Biology*, *27* (4), 407-421. Zugriff auf http://www.ncbi.nlm.nih.gov/pubmed/10942348

Chambers, J. M., Cleveland, W. S., Kleiner, B. & Tukey, P. A. (1983). *Graphical methods for data analysis*. New York, USA: The Wadsworth Statistics/Probability Series.

Chen, S. H. (2007). Sex differences in frequency and intensity in reading and voice range profiles for Taiwanese adult speakers. *Folia phoniatrica et logopaedica*, *59*, 1-9. Zugriff auf http://www.karger.com/Article/FullText/96545 doi: DOI:10.1159/000096545

Chen, S. H., Liu, H., Xu, Y. & Larson, C. (2007). Voice F0 responses to pitch-shifted voice feedback during English speech. *The Journal of the Acoustical Society of America*, *121*, 1157-1163. Zugriff auf http://dx.doi.org/10.1121/1.2404624

Childers, D. G. (1978). *Modern spectrum analysis* (Bd. 1). New York, USA: IEEE Computer Society Press.

Childers, D. G. & Wong, C. F. (1994). Measuring and modeling vocal source-tract interaction. *IEEE transactions on bio-medical engineering*, *41* (7), 663-671. Zugriff auf http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=301733&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D301733

Chistovich, L. A. & Lublinskaya, V. V. (1979). The 'center of gravity' effect in vowel spectra and critical distance between the formants: Psychoacoustical study of the perception of vowel-like stimuli. *Hearing research*, *1* (3), 185-195. Zugriff auf http://www.sciencedirect.com/science/article/pii/0378595579900121

Chládková, K., Boersma, P. & Podlipský, V. J. (2009). On-Line Formant Shifting as a Function of F0. In *Proceedings of INTERSPEECH 2009* (S. 464-467). Brighton, Vereinigtes Königreich. Zugriff auf http://www.fon.hum.uva.nl/paul/papers/IS090423.PDF

Chuang, C. & Wang, W. (1978). Psychophysical pitch biases related to vowel quality, intensity difference, and sequential order. *The Journal of the Acoustical Society of America*, *64* (4), 1004-1014. Zugriff auf http://dx.doi.org/10.1121/1.382083

Cohen, J. R., Crystal, T. H., House, A. S. & Neuburg, E. P. (1980). Weighty voices and shaky evidence: A critique. *Journal of the Acoustical Society of America*, *68* (6), 1884-1886. Zugriff auf http://dx.doi.org/10.1121/1.385178

Cohen, T. & Gitman, L. (1959). Oral complaints and taste perception in the aged. *Journal of Gerontology*, *14* (3), 294. Zugriff auf http://geronj.oxfordjournals.org/content/14/3/294.extract

Coleman, J., Grabe, E. & Braun, B. (2002). *Larynx movements and intonation in whispered speech.* Zugriff auf http://www.phon.ox.ac.uk/files/pdfs/project_larynx_summary.pdf

Connell, B. (2002). Tone languages and the universality of intrinsic F0: evidence from Africa. *Journal of Phonetics*, *30* (1), 101-129. Zugriff auf http://www.sciencedirect.com/science/article/pii/S0095447001901561

Corso, J. F. (1975). Sensory processes in man during maturity and senescence. In B. Odry (Hrsg.), *Neurobiology of aging.* New York, USA: Plenum Press.

da Silva, P. T., Master, S., Andreoni, S., Pontes, P. & Ramos, L. R. (2010). Acoustic and Long-Term Average Spectrum Measures to Detect Vocal Aging in Women. *Journal*

of Voice, *25* (4), 411-419.

Decoster, W. & Debruyne, F. (1997). The ageing voice: changes in fundamental frequency, waveform stability and spectrum. *Acta Oto-rhino-laryngologica Belgica*, *51* (2), 105-112.

Decoster, W. & Debruyne, F. (2000). Longitudinal voice changes: facts and interpretation. *Journal of Voice*, *14* (2), 184-193.

de Jong, K. (1997). Labiovelar compensation in back vowels. *The Journal of the Acoustical Society of America*, *101* (4), 2221. Zugriff auf http://dx.doi.org/10.1121/1.418206

Delattre, P., Liberman, A. M., Cooper, F. S. & Gerstman, L. J. (1952). An experimental study of the acoustic determinants of vowel color; observations on one-and two-formant vowels synthesized from spectrographic patterns. *Word*, *8*, 195-210.

De Pinto, O. & Hollien, H. (1982). Speaking fundamental frequency characteristics of Australian women: Then and now. *Journal of Phonetics*, *10*, 367-375.

Detweiler, R. F. (1994). An investigation of the laryngeal system as the resonance source of the singer's formant. *Journal of Voice*, *8* (4), 303-313. Zugriff auf http://dx.doi.org/10.1016/S0892-1997(05)80278-4

Di Benedetto, M.-G. (1987). On vowel height: Acoustic and perceptual representation by the fundamental and the first formant frequency. In *Proceedings of the XIth International Conference on Phonetic Sciences* (Bd. 5, S. 198-201). Tallinn, SU. Zugriff auf http://acts.ing.uniroma1.it/Papers/C10-DiBenedetto_al-ICPhS87.pdf

Di Benedetto, M. G. (1994). Acoustic and perceptual evidence of a complex relation between F_1 and F_0 in determining vowel height. *Journal of Phonetics*, *22* (3), 205-224. Zugriff auf http://psycnet.apa.org/psycinfo/1995-23980-001

Di Benedetto, M. G. (2003). Vowels: a revisit. *Studi in onore di Franco Ferrero*, 143-148. Zugriff auf http://newyork.diet.uniroma1.it/Papers/C43-DiBenedetto_al-MCSCI01.pdf

Diehl, R. L. & Kluender, K. R. (1989). On the Objects of Speech Perception. *Ecological Psychology*, *1* (2), 121-144. Zugriff auf http://www.tandfonline.com/doi/abs/10.1207/s15326969eco0102_2

Diehl, R. L., Lindblom, B., Hoemeke, K. A. & Fahey, R. P. (1996). On explaining certain male-female differences in the phonetic realization of vowel categories. *Journal of Phonetics*, *24* (2), 187-208. Zugriff auf http://dx.doi.org/10.1006/jpho.1996.0011

Disner, S. F. (1980). Evaluation of vowel normalization procedures. *The Journal of the Acoustical Society of America*, *67* (1), 253-261. Zugriff auf http://dx.doi.org/10.1121/1.383734

Donath, T. M., Natke, U. & Kalveram, K. T. (2002). Effects of frequency-shifted auditory feedback on voice F0 contours in syllables. *The Journal of the Acoustical Society of America*, *111*, 357-366. Zugriff auf http://dx.doi.org/10.1121/1.1424870

Drager, K. (2011). Speaker age and vowel perception. *Language and Speech*, *54* (1), 99-121. Zugriff auf http://las.sagepub.com/content/54/1/99.short

Draxler, C. (2011). Percy - an HTML5 framework for media rich web experiments on mobile

devices. In *Proceedings of the 12th Annual Conference of the International Speech Communication Association (Interspeech2011)* (S. 3339-3340). Florenz, Italien.

Dyhr, N. (1990). The activity of the cricothyroid muscle and the intrinsic fundamental frequency in Danish vowels. *Phonetica*, *47* (3-4), 141-154. Zugriff auf http://www.karger.com/Article/Abstract/261859

Eckert, P. (1997). Age as a Sociolinguistic Variable. In F. Coulmas (Hrsg.), *The handbook of sociolinguistics* (Bd. 4, S. 151-167). Oxford, Vereinigtes Königreich: Wiley-Blackwell.

Edwards, J. (1985). Contextual effects on lingual-mandibular coordination. *The Journal of the Acoustical Society of America*, *78* (6), 1944-1948. Zugriff auf http://dx.doi.org/10.1121/1.392650

Eklund, I. & Traunmüller, H. (1997). Comparative Study of Male and Female Whispered and Phonated Versions of the Long Vowels of Swedish. *Phonetica*, *54* (1), 1-21.

Endres, W., Bambach, W. & Flosser, G. (1971). Voice spectrograms as a function of age, voice disguise, and voice imitation. *Journal of the Acoustical Society of America*, *49* (6B), 1842-1848. Zugriff auf http://link.aip.org/link/?JAS/49/1842/1

Eriksson, A. & Traunmüller, H. (2002). Perception of vocal effort and distance from the speaker on the basis of vowel utterances. *Perception & psychophysics*, *64* (1), 131-139. Zugriff auf http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=11916296

Evans, S., Neave, N., Wakelin, D. & Hamilton, C. (2008). The relationship between testosterone and vocal frequencies in human males. *Physiology and Behavior*, *93* (4-5), 783-788. Zugriff auf http://www.sciencedirect.com/science/article/pii/S0031938407004775

Ewan, W. & Ohala, J. (1979). Can intrinsic vowel F0 be explained by source/tract coupling? *The Journal of the Acoustical Society of America*, *66* (2), 358-362. Zugriff auf http://dx.doi.org/10.1121/1.383669

Fahey, R. P. & Diehl, R. L. (1996). The missing fundamental in vowel height perception. *Attention, Perception, & Psychophysics*, *58* (5), 725-733. Zugriff auf http://link.springer.com/article/10.3758/BF03213105#page-1

Fahey, R. P., Diehl, R. L. & Traunmüller, H. (1996). Perception of back vowels: effects of varying F1 - F0 Bark distance. *The Journal of the Acoustical Society of America*, *99* (4), 2350-2357. Zugriff auf http://view.ncbi.nlm.nih.gov/pubmed/8730081

Fairbanks, G. (1955). Selective Vocal Effects Of Delayed Auditory Feedback. *Journal of Speech and Hearing Disorders*, *20* (4), 333-346. Zugriff auf http://psycnet.apa.org/index.cfm?fa=search.displayRecord&UID=1956-07109-001

Fant, G. (1975). Non-uniform vowel normalization. In *Speech transmission laboratory quarterly progress and status report* (Bd. 16, S. 1-19). Stockholm, Schweden: KTH. Zugriff auf http://www.speech.kth.se/prod/publications/files/qpsr/1975/1975_16_2-3_001-019.pdf

Fant, G., Carlson, R. & Granström, B. (1974). The [e]-[i] ambiguity. *Proceedings of the Speech Communication Seminar, Stockholm* (4), 117-121.

Ferreri, G. (1959). Senescence of the larynx. *Ital Gen Rev Otorhinolaryngol*, *1*, 640-709.

Fischer-Jørgensen, E. (1967). Phonetic analysis of breathy (murmured) vowels in Gujarati. *Indian linguistics*, *28*, 71-139.

Fischer-Jørgensen, E. (1990). Intrinsic F0 in Tense and Lax Vowels with Special Reference to German. *Phonetica*, *47* (3-4), 99-140.

Fitch, W. T. & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, *106* (3), 1511-1522. Zugriff auf `http://dx.doi.org/10.1121/1.427148`

Flanagan, J. (1965). *Speech analysis, synthesis and perception* (Bd. 3). Berlin [u.a.]: Springer.

Flanagan, J. & Landgraf, L. (1968). Self-oscillating source for vocal-tract synthesizers. *IEEE Transactions on Audio and Electroacoustics*, *16* (1), 57-64.

Flege, J. E. & Fletcher, S. G. (1988). Compensating for a bite block in/s/and/t/production: palatographic, acoustic, and perceptual data. *The Journal of the Acoustical Society of America*, *83* (1), 212-228. Zugriff auf `http://dx.doi.org/10.1121/1.396424`

Fletcher, H. (1940). Auditory patterns. *Reviews of Modern Physics*, *12* (1), 47-65.

Flügel, C. & Rohen, J. W. (1991). The craniofacial proportions and laryngeal position in monkeys and man of different ages. A morphometric study based on CT-scans and radiographs. *Mechanisms of Ageing and Development* (61), 65-83. Zugriff auf `http://dx.doi.org/10.1016/0047-6374(91)90007-M`

Folkins, J. W. & Abbs, J. H. (1975). Lip and jaw motor control during speech: Responses to resistive loading of the jaw. *Journal of Speech and Hearing Research*, *18* (1), 207-220. Zugriff auf `http://www.ncbi.nlm.nih.gov/pubmed/1127904?dopt=Abstract`

Fowler, C. & Brown, J. (1997). Intrinsic F0 differences in spoken and sung vowels and their perception by listeners. *Attention, Perception, & Psychophysics*, *59* (5), 729-738. Zugriff auf `http://dx.doi.org/10.3758/BF03206019`

Fox, J. & Weisberg, S. (2011). *An R Companion to Applied Regression* (2. Aufl.). Thousand Oaks, CA., USA: Sage. Zugriff auf `http://socserv.socsci.mcmaster.ca/jfox/Books/Companion`

Fredberg, J. J., Wohl, M. E., Glass, G. M. & Dorkin, H. L. (1980). Airway area by acoustic reflections measured at the mouth. *Journal of Applied Physiology*, *48* (5), 749-758. Zugriff auf `http://www.ncbi.nlm.nih.gov/pubmed/7451282`

Frøkjær-Jensen, B. (1966). Changes in formant frequencies and formant levels at high voice effort. In *Annual Report of the Institute of Phonetics I* (S. 47-55). Kopenhagen, Dänemark: Institute of Phonetics, Universität Kopenhagen.

Fujisaki, H. & Kawashima, T. (1968). The roles of pitch and higher formants in the perception of vowels. *IEEE Transactions on Audio and Electroacoustics*, *16* (1), 73-77.

Garnier, M., Henrich, N., Smith, J. & Wolfe, J. (2010). The tuning of vocal resonances and the upper limit to the high soprano range. In *Proceedings of the International Symposium on Music Acoustics* (S. 1-6). Sydney & Katoomba, Australien.

Garnier, M., Wolfe, J., Henrich, N. & Smith, J. (2008). Interrelationship between vocal effort and vocal tract acoustics: a pilot study. *Proceedings of the Ninth Annual Conference of the International Speech Communication Association*.

Gerstman, L. (1968). Classification of self-normalized vowels. *IEEE Transactions on Audio and Electroacoustics*, *16* (1), 78-80.

Geumann, A. (2001a). Invariance and variability in articulation and acoustics of natural perturbed speech. In P. Hoole (Hrsg.), *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München* (Bd. 38, S. 265-393). München: Institut für Phonetik und Sprachliche Kommunikation.

Geumann, A. (2001b). Vocal Intensity: Acoustic and Articulatory Correlates. In *Proceedings of the 4th International Speech Motor Conference* (S. 70–73). Nijmegen, Niederlande.

Geumann, A., Kroos, C. & Tillmann, H. G. (1999). Are there compensatory effects in natural speech? *Proceedings of the 14th International Conference on Phonetic Sciences, San Francisco*, 399-402.

Ghahramani, Z., Wolpert, D. M. & Jordan, M. I. (1996). Generalization to local remappings of the visuomotor coordinate transformation. *The Journal of Neuroscience*, *16* (21), 7085-7096.

Ghosh, S. S., Matthies, M. L., Maas, E., Hanson, A., Tiede, M., Ménard, L., ... Perkell, J. S. (2010). An investigation of the relation between sibilant production and somatosensory and auditory acuity. *The Journal of the Acoustical Society of America*, *128* (5), 3079-3087. Zugriff auf `http://dx.doi.org/10.1121/1.3493430`

Golden, R. (1966). Improving Naturalness and Intelligibility of Helium-Oxygen Speech, Using Vocoder Techniques. *The Journal of the Acoustical Society of America*, *40* (3), 621-624. Zugriff auf `http://dx.doi.org/10.1121/1.1910127`

Goldstein, U. G. (1980). *An articulatory model for the vocal tracts of growing children* (Dissertation). Zugriff auf `http://18.7.29.232/bitstream/handle/1721.1/22386/Goldstein_Ursula_ScD_1980.pdf?sequence=1`

González, J. (2004). Formant frequencies and body size of speaker: a weak relationship in adult humans. *Journal of Phonetics*, *32* (2), 277-287. Zugriff auf `http://www.sciencedirect.com/science/article/B6WKT-4B3NM6D-2/2/7e1307b6b1eab7f72c18fe98b173f573`

Gottfried, T. & Chew, S. (1986). Intelligibility of vowels sung by a countertenor. *The Journal of the Acoustical Society of America*, *79* (1), 124-130. Zugriff auf `http://dx.doi.org/10.1121/1.393635`

Graddol, D. & Swann, J. (1983). Speaking fundamental frequency: some physical and social correlates. *Language & Speech*, *26* (4), 351-366. Zugriff auf `http://las.sagepub.com/content/26/4/351.short`

Guenther, F. H. (1994). A neural network model of speech acquisition and motor equivalent speech production. *Biological Cybernetics*, *72* (1), 43-53.

Guenther, F. H. (1995). Speech Sound Acquisition, Coarticulation, and Rate Effects in a Neural Network Model of Speech Production. *Psychological Review*, *102*, 594-621.

Guenther, F. H., Ghosh, S. S., Nieto-Castanon, A. & Tourville, J. A. (2006). A neural model of speech production. In J. Harrington & M. Tabain (Hrsg.), *Speech Production* (S. 27-39). New York, USA: Psychology Press.

Guenther, F. H., Ghosh, S. S. & Tourville, J. A. (2006). Neural modeling and imaging

of the cortical interactions underlying syllable production. *Brain and Language*, *96* (3), 280-301. Zugriff auf http://www.sciencedirect.com/science/article/pii/S0093934X0500115X

Guenther, F. H., Hampson, M. & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review*, *105* (4), 611-651.

Gugatschka, M., Kiesler, K., Obermayer-Pietsch, B., Schoekler, B., Schmid, C., Groselj-Strele, A. & Friedrich, G. (2010). Sex hormones and the elderly male voice. *Journal of Voice*, *24* (3), 369-373. Zugriff auf http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=19185460

Guimarães, I. & Abberton, E. (2005). Fundamental frequency in speakers of portuguese for different voice samples. *Journal of Voice*, *19* (4), 592-606.

Hanson, H. M. (1995). Individual variations in glottal characteristics of female speakers. *Acoustics, Speech, and Signal Processing, 1995. ICASSP-95*, *1*.

Hanson, H. M. (1997). Glottal characteristics of female speakers: acoustic correlates. *The Journal of the Acoustical Society of America*, *101* (1), 466-481.

Hanson, H. M. & Chuang, E. S. (1999). Glottal characteristics of male speakers: Acoustic correlates and comparison with female data. *The Journal of the Acoustical Society of America*, *106* (2), 1064-1077. Zugriff auf http://link.aip.org/link/?JAS/106/1064/1/http://dx.doi.org/10.1121/1.427116

Harnsberger, J. D., Wright, R. & Pisoni, D. B. (2008). A new method for eliciting three speaking styles in the laboratory. *Speech Communication*, *50* (4), 323-336. Zugriff auf http://www.sciencedirect.com/science/article/pii/S0167639307001859

Harrington, J. (2006). An acoustic analysis of 'happy-tensing' in the Queen's Christmas broadcasts. *Journal of Phonetics*, *34* (4), 439-457. Zugriff auf http://www.sciencedirect.com/science/article/pii/S0095447005000513

Harrington, J. (2007). Evidence for a relationship between synchronic variability and diachronic change in the Queen's annual Christmas broadcasts. In J. Cole & J. I. Hualde (Hrsg.), *Laboratory phonology 9* (Bde. 4,3, S. 125-143). Berlin: Mouton de Gruyter.

Harrington, J. (2010). *Phonetic Analysis of Speech Corpora*. Chichester: Wiley-Blackwell.

Harrington, J. (2011). *Notes and verification of the phoc() function*. Zugriff auf http://www.phonetik.uni-muenchen.de/~jmh/lehre/sem/ss11/statfort/posthoc.pdf

Harrington, J. & IPS LMU Muenchen & IPDS CAU Kiel. (2011). *emu: Interface to the Emu Speech Database System*. Zugriff auf http://CRAN.R-project.org/package=emu

Harrington, J., Kleber, F. & Reubold, U. (2008). Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study. *The Journal of the Acoustical Society of America*, *123* (5), 2825-2835. Zugriff auf http://scitation.aip.org/content/asa/journal/jasa/123/5/10.1121/1.2897042

Harrington, J., Palethorpe, S. & Watson, C. (2000b). Monophthongal vowel changes in Received Pronunciation: an acoustic analysis of the Queen's Christmas broadcasts. *Journal of the International Phonetic Association*, *30* (1/2), 63-78.

Harrington, J., Palethorpe, S. & Watson, C. I. (2000a). Does the Queen speak the Queen's

English? *Nature*, *408* (6815), 927-928. Zugriff auf http://dx.doi.org/10.1038/35050160

Harrington, J., Palethorpe, S. & Watson, C. L. (2007a). Age-related changes in fundamental frequency and formants: a longitudinal study of four speakers. In *Proceedings of INTERSPEECH 2007* (S. 2753-2756). Antwerpen, Belgien.

Harrington, J., Palethorpe, S. & Watson, C. L. (2007b). Deepening or lessening the divide between diphthongs? An analysis of the Queen's annual Christmas broadcasts. In J. Cole & J. I. Hualde (Hrsg.), *Laboratory phonology 9* (Bde. 4,3, S. 227-261). Berlin: Mouton de Gruyter.

Hawco, C. S., Jones, J. A., Ferretti, T. R. & Keough, D. (2009). ERP correlates of online monitoring of auditory feedback during vocalization. *Psychophysiology*, *46* (6), 1216-1225. Zugriff auf http://dx.doi.org/10.1111/j.1469-8986.2009.00875.x

Hawkins, S. & Midgley, J. (2005). Formant frequencies of RP monophthongs in four age groups of speakers. *Journal of the International Phonetic Association* (35), 183-199.

Hay, J. & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics*, *48* (4), 865-892.

Heid, S. J. G. G. (1998). Phonetische Variation: Untersuchungen anhand des PhonDat2-Korpus. In P. Hoole (Hrsg.), *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München 36* (S. 193-368). München.

Henrich, N., Smith, J. & Wolfe, J. (2011). Vocal tract resonances in singing: Strategies used by sopranos, altos, tenors, and baritones. *The Journal of the Acoustical Society of America*, *129*, 1024-1035.

Herries, J. (1974). *The elements of speech*. London: Scolar press.

Higgins, M. B., Netsell, R. & Schulte, L. (1998). Vowel-Related Differences in Laryngeal Articulatory and Phonatory Function. *Journal of Speech, Language & Hearing Research*, *41* (4), 712-724. Zugriff auf http://jslhr.pubs.asha.org/article.aspx?articleid=1780567

Higgins, M. B. & Saxman, J. H. (1989). Variations in vocal frequency perturbation across the menstrual cycle. *Journal of Voice*, *3* (3), 233-243. Zugriff auf http://www.sciencedirect.com/science/article/B7585-4G83R99-5/2/26a735f5e24e1eb74803001677698b91

Hillenbrand, J., Cleveland, R. A. & Erickson, R. L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech & Hearing Research*, *37*, 769-778.

Hillenbrand, J. & Gayvert, R. T. (1993). Vowel classification based on fundamental frequency and formant frequencies. *Journal of Speech & Hearing Research*, *36*, 694-700.

Hillenbrand, J., Getty, L. A., Clark, M. J. & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, *97* (5), 3099-3111.

Hirahara, T. & Kato, H. (1992). The Effect of F0 on Vowel Identification. In Y. Tohkura & E. Vatikiotis-Bateson (Hrsg.), *Speech Perception, Production and Linguistic Structure* (S. 88-111). Burke, VA, USA: IOS Press. Zugriff auf http://books.google.de/books?hl=de&lr=&id=eYrOR7VHnnQC&oi=fnd&pg=PA89&dq=Hirahara+%26+Kato+

1992&ots=p6hajS_zSt&sig=F7QE_nnCR1owR_r0r4A1j4xwg0I

Hirano, M., Kurita, S. & Nakashima, T. (1983). Growth, development and aging of human vocal folds. *Vocal fold physiology: Contemporary research and clinical issues*, 22-43.

Hirano, M., Kurita, S. & Sakaguchi, S. (1988). Vocal fold tissue of a 104-year-old lady. *Annual Bulletin RILP*, *22*, 1-5.

Hirano, M., Kurita, S. & Sakaguchi, S. (1989). Ageing of the vibratory tissue of human vocal folds. *Acta Oto-Laryngologica*, *107* (5-6), 428-433.

Hoemeke, K. A. & Diehl, R. L. (1994). Perception of vowel height: the role of F1-F0 distance. *The Journal of the Acoustical Society of America*, *96* (2), 661-674. Zugriff auf http://view.ncbi.nlm.nih.gov/pubmed/7930066

Hoit, J. D. & Hixon, T. J. (1987). Age and speech breathing. *Journal of Speech and Hearing Research*, *30* (3), 351-366. Zugriff auf http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=3669642

Hoit, J. D. & Hixon, T. J. (1992). Age and laryngeal airway resistance during vowel production in women. *Journal of Speech and Hearing Research*, *35* (2), 309-313. Zugriff auf http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=1573871

Hollien, H. (1987). "Old voices": What do we really know about them? *Journal of Voice*, *1* (1), 2-17. Zugriff auf http://www.sciencedirect.com/science/article/pii/S0892199787800188

Hollien, H., Hollien, P. A. & de Jong, G. (1997). Effects of three parameters on speaking fundamental frequency. *The Journal of the Acoustical Society of America*, *102* (5), 2984-2992. Zugriff auf http://dx.doi.org/10.1121/1.420353

Hollien, H. & Shipp, T. (1972). Speaking fundamental frequency and chronologic age in males. *Journal of Speech and Hearing Research*, *15* (1), 155-159. Zugriff auf http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=5012800

Holmes, J. & Holmes, W. (2001). *Speech synthesis and recognition* (2. Aufl.). London [u.a.]: Routledge.

Holt, L., Lotto, A. & Kluender, K. (2001). Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement? *The Journal of the Acoustical Society of America*, *109* (2), 764-774. Zugriff auf http://dx.doi.org/10.1121/1.1339825

Honda, K. (2004). Physiological factors causing tonal characteristics of speech: from global to local prosody. In *Proceedings of Speech Prosody 2004.* Nara, Japan.

Honda, K. & Fujimura, O. (1991). Intrinsic vowel f0 and phrase-final f0 lowering: phonological vs. biological explanations. In *Vocal fold physiology* (Bd. 3). San Diego: Singular Publishing Group.

Honda, K., Hirai, H., Masaki, S. & Shimada, Y. (1999). Role of vertical larynx movement and cervical lordosis in F0 control. *Language and speech*, *42* (4), 401-411.

Honda, M., Fujino, A. & Kaburagi, T. (2002). Compensatory responses of articulators to unexpected perturbation of the palate shape. *Journal of Phonetics*, *30* (3), 281-302.

Honjo, I. & Isshiki, N. (1980). Laryngoscopic and Voice Characteristics of Aged Persons. *Archives of otolaryngology*, *106* (3), 149-150. Zugriff auf `http://archotol.ama-assn.org/cgi/content/abstract/106/3/149`

Hoole, P. (1987). Bite-block speech in the absence of oral sensibility. In *Proceedings of the 11th International Congress on Phonetic Sciences* (Bd. 4). Tallinn, SU.

Hoole, P. (2006). *Experimental studies of laryngeal articulation* (Habilitation, Ludwig-Maximilians-Universität, München). Zugriff auf `http://www.phonetik.uni-muenchen.de/~hoole/pdf/habilemg_chap_all.pdf`

Hoole, P. & Honda, K. (2011). Automaticity vs. feature-enhancement in the control of segmental F0. In G. N. Clements & R. Ridouane (Hrsg.), *Where do phonological contrasts come from? Cognitive, physical and developmental bases of phonological features* (S. 131-171). Amsterdam, Niederlande: John Benjamins Publishing.

Hoole, P., Honda, K., Murano, E., Fuchs, S. & Pape, D. (2004). Cricothyroid activity in consonant voicing and vowel intrinsic pitch. In *Proceedings of the Conference on Voice Physiology and Biomechanics.* Marseille.

Hoole, P. & Kroos, C. (1998). Control of larynx height in vowel production. In *Proceedings of the 5th International Conference Spoken Language Processing* (Bd. 2, S. 531–534). Sidney, Australien.

Hoole, P. & Mooshammer, C. (2002). Articulatory analysis of the German vowel system. In P. Auer, P. Gilles & H. Spiekermann (Hrsg.), *Silbenschnitt und Tonakzente* (S. 129–152). Tübingen: Niemeyer.

Hothorn, T., Bretz, F. & Westfall, P. (2008). Simultaneous Inference in General Parametric Models. *Biometrical Journal*, *50* (3), 346–363.

Houde, J. F. & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, *279* (5354), 1213-1216.

Houde, J. F. & Jordan, M. I. (2002). Sensorimotor Adaptation of Speech I: Compensation and Adaptation. *Journal of Speech, Language and Hearing Research*, *45* (2), 295-310. Zugriff auf `http://jslhr.asha.org/cgi/content/abstract/45/2/295`

Huber, J. E. & Spruill, J. (2008). Age-related changes to speech breathing with increased vocal loudness. *Journal of speech, language, and hearing research : JSLHR*, *51* (3), 651-668. Zugriff auf `http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=18506042`

Hughes, M. O. & Abbs, J. H. (1976). Labial-mandibular coordination in the production of speech: Implications for the operation of motor equivalence. *Phonetica*, *33* (3), 199-221.

Imaizumi, S., Mori, K., Kiritani, S., Kawashima, R., Sugiura, M., Fukuda, H., ... Hatano, K. (1997). Vocal identification of speaker and emotion activates differerent brain regions. *Neuroreport*, *8* (12), 2809-2812.

Iseli, M., Shue, Y. L. & Alwan, A. (2006). Age-and gender-dependent analysis of voice source characteristics. In *Acoustics, Speech and Signal Processing, 2006. ICASSP Proceedings* (Bd. 1, S. 389-392). Toulouse, Frankreich.

Iseli, M., Shue, Y. L. & Alwan, A. (2007). Age, sex, and vowel dependencies of acoustic measures related to the voice source. *The Journal of the Acoustical Society of*

America, *121* (4), 2283-2295.

Ishizaka, K. & Flanagan, J. (1972). Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell SystemTechnical Journal* (51), 1233-1268.

Israel, H. (1968). Continuing growth in the human cranial skeleton. *Archives of Oral Biology*, *13*, 133-137.

Israel, H. (1973). Age factor and the pattern of change in craniofacial structures. *American Journal of Physical Anthropology*, *39*, 111-128.

Iwarsson, J. & Sundberg, J. (1998). Effects of lung volume on vertical larynx position during phonation. *Journal of Voice*, *12* (2), 159-165. Zugriff auf `http://www.sciencedirect.com/science/article/B7585-4GCP29P-5/2/b66066019de5846e8c468ca8b634f2e3`

Jiang, D.-N., Tao, J.-H. & Cai, L.-H. (2002). Voice quality analysis under the pitch effect. In *Proceedings of the International Symposium on Chinese Spoken Language Processing* (S. o. S.). Taipei, Taiwan. Zugriff am 2013-05-06 auf `http://www.isca-speech.org/archive_open/iscslp2002/clp2_109.html`

Johnson, K. (1989a). F0 normalization and talker variability. *The Journal of the Acoustical Society of America*, *85* (S1), S51. Zugriff auf `http://link.aip.org/link/?JAS/85/S51/3/http://dx.doi.org/10.1121/1.2027009`

Johnson, K. (1989b). Higher formant normalization results from auditory integration of F2 and F3. *Attention, Perception, & Psychophysics*, *46* (2), 174-180.

Johnson, K. (1990). The role of perceived speaker identity in F0 normalization of vowels. *The Journal of the Acoustical Society of America*, *88* (2), 642-654. Zugriff auf `http://link.aip.org/link/?JAS/88/642/1/http://dx.doi.org/10.1121/1.399767`

Johnson, K. (2005). Speaker normalization in speech perception. In D. B. Pisoni & R. E. Remez (Hrsg.), *The handbook of speech perception* (S. 363-389). Oxford, Vereinigtes Königreich: Blackwell.

Johnson, K., Strand, E. A. & D'Imperio, M. (1999). Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics*, *27* (4), 359-384.

Joliveau, E., Smith, J. & Wolfe, J. (2004a). Acoustics: Tuning of vocal tract resonance by sopranos. *Nature*, *427*, 116.

Joliveau, E., Smith, J. & Wolfe, J. (2004b). Vocal tract resonances in singing: The soprano voice. *The Journal of the Acoustical Society of America*, *116*, 2434-2439.

Jones, D. & Roach, P. (2009). *Cambridge English pronouncing dictionary: spoken pronunciation for every word - now American English too* (17[4] Aufl.). Cambridge: Cambridge Univ. Press.

Jones, J. A. & Keough, D. (2008). Auditory-motor mapping for pitch control in singers and nonsingers. *Experimental Brain Research*, *190* (3), 279-287. Zugriff auf `http://dx.doi.org/10.1007/s00221-008-1473-y`

Jones, J. A. & Munhall, K. G. (2000). Perceptual calibration of F0 production: Evidence from feedback perturbation. *The Journal of the Acoustical Society of America*, *108* (3), 1246-1251. Zugriff auf `http://link.aip.org/link/?JAS/108/1246/1/http://dx.doi.org/10.1121/1.1288414`

Jones, J. A. & Munhall, K. G. (2003). Learning to produce speech with an altered vocal tract: The role of auditory feedback. *The Journal of the Acoustical Society of America*, *113* (1), 532-543. Zugriff auf `http://link.aip.org/link/?JAS/113/532/1`

Jones, J. A. & Munhall, K. G. (2005). Remapping Auditory-Motor Representations in Voice Production. *Current Biology*, *15* (19), 1768-1772. Zugriff auf `http://www.sciencedirect.com/science/article/pii/S0960982205010225`

Junqua, J.-C. (1996). The influence of acoustics on speech production: A noise-induced stress phenomenon known as the Lombard reflex. *Speech Communication*, *20* (1-2), 13-22. Zugriff auf `http://www.sciencedirect.com/science/article/pii/S0167639396000416`

Kahane, J. (1980). Age related histological changes in the human male and female laryngeal cartilages: Biological and functional implications. In *Transcripts of the Ninth Symposium: Care of the Professional Voice.* New York, USA: The Voice Foundation.

Katseff, S. E., Houde, J. F. & Johnson, K. (2010). Auditory feedback shifts in one formant cause multi-formant responses. *The Journal of the Acoustical Society of America*, *127* (3), 1955. Zugriff auf `http://link.aip.org/link/?JAS/127/1955/3/http://dx.doi.org/10.1121/1.3384960`

Katz, W. F. & Assmann, P. F. (2001). Identification of children's and adults' vowels: intrinsic fundamental frequency, fundamental frequency dynamics, and presence of voicing. *Journal of Phonetics*, *29* (1), 23-51. Zugriff auf `http://www.sciencedirect.com/science/article/B6WKT-457CJ0D-C/2/69db54426476d0aa1fbdd0a840b480f6`

Kawahara, H. (1993). Transformed auditory feedback: Effects of fundamental frequency perturbation. *The Journal of the Acoustical Society of America*, *94*, 1883.

Kawahara, H. & Irino, T. (2005). Underlying principles of a high-quality speech manipulation system STRAIGHT and its application to speech segregation. In *Speech separation by humans and machines* (S. 167-180). New York, USA: Springer.

Kawahara, H., Masuda-Katsuse, I. & de Cheveigné, A. (1999). Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds. *Speech Communication*, *27* (3-4), 187-207. Zugriff auf `http://www.sciencedirect.com/science/article/B6V1C-3W49BY3-4/2/0aeba8270135ad58dc12c8b421b73c40`

Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T. & Banno, H. (2008). TANDEM-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. In *Proceedings of ICASSP* (S. 3933-3936). Las Vegas, USA.

Kawahara, H., Nisimura, R., Irino, T., Morise, M., Takahashi, T. & Banno, H. (2009). Temporally variable multi-aspect auditory morphing enabling extrapolation without objective and perceptual breakdown. In *Proceedings of ICASSP* (S. 19-24). Taipei, Taiwan.

Kawahara, H., Takahashi, T., Morise, M. & Banno, H. (2009). Development of exploratory research tools based on TANDEM-STRAIGHT. In *Proceedings of APSIPA ASC 2009: Asia-Pacific Signal and Information Processing Association, 2009 An-*

nual Summit and Conference (S. 111-120). Sapporo, Japan: Asia-Pacific Signal and Information Processing Association.

Kelso, J. A. S. & Tuller, B. (1983). "Compensatory Articulation" Under Conditions of Reduced Afferent Information: A Dynamic Formulation. *Journal of Speech and Hearing Research*, *26* (2), 217-224. Zugriff auf `http://jslhr.asha.org/cgi/content/abstract/26/2/217`

Kelso, J. A. S., Tuller, B., Vatikiotis-Bateson, E. & Fowler, C. A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance*, *10* (6), 812-832.

Kent, R. D. (1976). Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies. *Journal of Speech and Hearing Research*, *19* (3), 421-447.

Kent, R. D. (1993). Vocal tract acoustics. *Journal of Voice*, *7* (2), 97-117. Zugriff auf `http://www.sciencedirect.com/science/article/B7585-4JDBM78-1/2/9a0e81dc14fbfff7506a7079f111e9b7`

King, A., Ashby, J. & Nelson, C. (2001). Effects of Testosterone Replacement on a Male Professional Singer. *Journal of Voice*, *15* (4), 553-557.

Kingston, J. (1992). The Phonetics and Phonology of Perceptually Motivated Articulatory Covariation. *Language and Speech*, *35* (1-2), 99-113. Zugriff auf `http://las.sagepub.com/content/35/1-2/99.abstract`

Kipp, A. (1999). *Automatische Segmentierung und Etikettierung von Spontansprache*. Aachen: Shaker.

Klatt, D. H. (1973). Discrimination of fundamental frequency contours in synthetic speech: implications for models of pitch perception. *The Journal of the Acoustical Society of America*, *53* (1), 8-16.

Klatt, D. H. & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of the Acoustical Society of America*, *87* (2), 820-857.

Kochanski, G., Grabe, E., Coleman, J. & Rosner, B. (2005). Loudness predicts prominence: Fundamental frequency lends little. *The Journal of the Acoustical Society of America*, *118* (2), 1038-1054.

König, W. (1989). *Atlas zur Aussprache des Schriftdeutschen in der Bundesrepublik Deutschland*. München: Hueber.

König, W. (2009). *Kleiner bayerischer Sprachatlas* (3., korr. u. überarb. Aufl., Bd. 3328). München: Deutscher Taschenbuch-Verlag (DTV).

Künzel, H. J. (1989). How well does average fundamental frequency correlate with speaker height and weight? *Phonetica*, *46* (1-3), 117-125.

Labov, W., Yaeger, M. & Steiner, R. (1972). *A quantitative study of sound change in progress* (Bd. 1). US Regional Survey.

Ladefoged, P. (2005). *Vowels and consonants: An introduction to the sounds of languages* (2. Aufl.). Malden, MA [u.a.]: Blackwell.

Ladefoged, P. & Broadbent, D. E. (1957). Information conveyed by vowels. *The Journal of the Acoustical Society of America*, *29* (1), 98-104.

Ladefoged, P., Maddieson, I. & Jackson, M. (1988). Investigating phonation types in different languages. In O. Fujimura (Hrsg.), *Vocal physiology: Voice production, mechanisms and functions* (S. 297-317). New York, USA: Raven Press.

Ladefoged, P. & McKinney, N. P. (1963). Loudness, sound pressure, and subglottal pressure in speech. *The Journal of the Acoustical Society of America*, *35*, 454-460.

Lane, H., Perkell, J., Wozniak, J., Manzella, J., Guiod, P., Matthies, M., ... Vick, J. (1998). The effect of changes in hearing status on speech sound level and speech breathing: A study conducted with cochlear implant users and NF-2 patients. *The Journal of the Acoustical Society of America*, *104* (5), 3059-3069. Zugriff auf `http://link.aip.org/link/?JAS/104/3059/1`

Lane, H. & Tranel, B. (1971). The Lombard Sign and the Role of Hearing in Speech. *Journal of Speech and Hearing Research*, *14* (4), 677-709. Zugriff auf `http://jslhr.asha.org/cgi/content/abstract/14/4/677`

Lane, H. & Webster, J. (1991). Speech deterioration in postlingually deafened adults. *The Journal of the Acoustical Society of America*, *89* (2), 859-866.

Lane, H., Wozniak, J., Matthies, M., Svirsky, M., Perkell, J., O'Connell, M. & Manzella, J. (1997). Changes in sound pressure and fundamental frequency contours following changes in hearing status. *The Journal of the Acoustical Society of America*, *101* (4), 2244-2252. Zugriff auf `http://dx.doi.org/10.1121/1.418245`

Larson, C., Altman, K., Liu, H. & Hain, T. (2008). Interactions between auditory and somatosensory feedback for voice F0 control. *Experimental Brain Research*, *187* (4), 613-621. Zugriff auf `http://dx.doi.org/10.1007/s00221-008-1330-z`

Larson, C., Burnett, T. A., Kiran, S. & Hain, T. C. (2000). Effects of pitch-shift velocity on voice F0 responses. *The Journal of the Acoustical Society of America*, *107*, 559.

Larsson, H. & Hertegård, S. (2008). Vocal Fold Dimensions in Professional Opera Singers as Measured by Means of Laser Triangulation. *Journal of Voice*, *22* (6), 734-739. Zugriff auf `http://www.sciencedirect.com/science/article/pii/S0892199707000148`

Lass, N. J. (1981). A reply to Cohen et al.'s "Weighty voices and shaky evidence: a critique" [J. Acoust. Soc. Am. 68, 1884–1886 (1980)]. *The Journal of the Acoustical Society of America*, *69* (4), 1204-1206. Zugriff auf `http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=7229204`

Lass, N. J., Beverly, A. S., Nicosia, D. K. & La Simpson. (1978). An investigation of speaker height and weight identification by means of direct estimations. *Journal of Phonetics*, *6*, 69-76.

Lass, N. J. & Davis, M. (1976). An investigation of speaker height and weight identification. *The Journal of the Acoustical Society of America*, *60*, 700-703.

Lass, N. J., Dicola, G. A., Beverly, A. S., Barbera, C., Henry, K. G. & Badali, M. K. (1979). The effect of phonetic complexity on speaker height and weight identification. *Language and Speech*, *22* (4), 297-309.

Lass, N. J., Phillips, J. K. & Bruchey, C. A. (1980). The effect of filtered speech on speaker height and weight identification. *Journal of Phonetics*, *8*, 91-100.

Laver, J. & Trudgill, K. (1979). Phonetic and linguistic markers in speech. In H. Giles & K. R. Scherer (Hrsg.), *Social Markers in speech.* Cambridge, Vereinigtes Königreich: Cambridge University Press.

Lawrence, M. A. (2011). *ez: Easy analysis and visualization of factorial experiments.* Zugriff auf `http://CRAN.R-project.org/package=ez`

Lee, S., Potamianos, A. & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*, *105* (3), 1455-1468. Zugriff auf `http://link.aip.org/link/?JAS/105/1455/1`

Lehiste, I. & Meltzer, D. (1973). Vowel and speaker identification in natural and synthetic speech. *Language and Speech*, *16* (4), 356-364.

Lieberman, D. E., McCarthy, R. C., Hiiemae, K. M. & Palmer, J. B. (2001). Ontogeny of postnatal hyoid and larynx descent in humans. *Archives of Oral Biology*, *46* (2), 117-128. Zugriff auf `http://www.sciencedirect.com/science/article/B6T4J-4233N49-3/2/999e47de46d57a4c94607857ddcf4b74`

Lieberman, P. (1970). A study of prosodic features. *Haskins Lab. Status Rep. Speech Res*, *23*, 179-208.

Liénard, J.-S. & Di Benedetto, M.-G. (1999). Effect of vocal effort on spectral properties of vowels. *The Journal of the Acoustical Society of America*, *106* (1), 411-422. Zugriff auf `http://link.aip.org/link/?JAS/106/411/1`

Lindblom, B., Lubker, J. & Gay, T. (1977). Formant frequencies of some fixed mandible vowels and a model of speech motor programming by predictive simulation. *The Journal of the Acoustical Society of America*, *62*, S15.

Lindblom, B., Lubker, J. & McAllister, R. (1977). Compensatory articulation and the modeling of normal speech production behavior. In *Articulatory modeling and phonetics (Proceedings from Symposium at Grenoble, GALF).*

Lindblom, B. & Schulman, R. (1982). The target theory of speech production in the light of mandibular dynamics. In *Proceedings of the Autumn Conference of the British Institute of Acoustics* (S. A2.1-A2-5). Bournemouth, Vereinigtes Königreich.

Lindblom, B. & Sundberg, J. (1971). Acoustical consequences of lip, tongue, jaw, and larynx movement. *The Journal of the Acoustical Society of America*, *50* (4), 1166-1179.

Linville, S. E. (1987a). Acoustic-perceptual studies of aging voice in women. *Journal of Voice*, *1* (1), 44-48. Zugriff auf `http://linkinghub.elsevier.com/retrieve/pii/S0892199787800231?showall=true`

Linville, S. E. (1987b). Maximum phonational frequency range capabilities of women's voices with advancing age. *Folia phoniatrica*, *39* (6), 297-301.

Linville, S. E. (1996). The sound of senescence. *Journal of Voice*, *10* (2), 190-200.

Linville, S. E. (2001). *Vocal aging.* San Diego, Calif., USA: Singular Thomson Learning.

Linville, S. E. (2002). Source characteristics of aged voice assessed from long-term average spectra. *Journal of Voice*, *16* (4), 472-479.

Linville, S. E. & Fisher, H. B. (1985a). Acoustic characteristics of perceived versus actual vocal age in controlled phonation by adult females. *The Journal of the Acoustical*

Society of America, *78* (1 Pt 1), 40-48.

Linville, S. E. & Fisher, H. B. (1985b). Acoustic characteristics of women's voices with advancing age. *Journal of Gerontology*, *40* (3), 324-330.

Linville, S. E. & Rens, J. (2001). Vocal tract resonance analysis of aging voice using long-term average spectra. *Journal of Voice*, *15* (3), 323-330.

Liu, H., Auger, J. & Larson, C. (2010). Voice fundamental frequency modulates vocal response to pitch perturbations during English speech. *The Journal of the Acoustical Society of America*, *127* (1), EL1-5.

Liu, H., Russo, N. M. & Larson, C. (2010). Age-related differences in vocal responses to pitch feedback perturbations: A preliminary study. *The Journal of the Acoustical Society of America*, *127* (2), 1042-1046. Zugriff auf `http://link.aip.org/link/?JAS/127/1042/1/http://dx.doi.org/10.1121/1.3273880`

Liu, H., Zhang, Q., Xu, Y. & Larson, C. (2007). Compensatory responses to loudness-shifted voice feedback during production of Mandarin speech. *The Journal of the Acoustical Society of America*, *122*, 2405-2412.

Lobanov, B. (1971). Classification of Russian Vowels Spoken by Different Speakers. *The Journal of the Acoustical Society of America*, *49* (2B), 606-608. Zugriff auf `http://dx.doi.org/10.1121/1.1912396`

Lubker, J., McAllister, R. & Lindblom, B. (1977). Vowel fundamental frequency and tongue height. *The Journal of the Acoustical Society of America*, *62* (S1), S16. Zugriff auf `http://dx.doi.org/10.1121/1.2016048`

MacDonald, E. N., Goldberg, R. & Munhall, K. G. (2010). Compensations in response to real-time formant perturbations of different magnitudes. *The Journal of the Acoustical Society of America*, *127* (2), 1059-1068. Zugriff auf `http://link.aip.org/link/?JAS/127/1059/1`

MacDonald, E. N., Purcell, D. W. & Munhall, K. G. (2011). Probing the independence of formant control using altered auditory feedback. *The Journal of the Acoustical Society of America*, *129* (2), 955-965. Zugriff auf `http://link.aip.org/link/?JAS/129/955/1`

Maeda, S. (1979). Un modèle articulatoire basé sur une étude acoustique. In *Actes 10èmes Journé d'Etudes sur la Parole* (S. 152-162). Grenoble, Frankreich.

Maeda, S. (1990). Compensatory articulation during speech: evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. In W. J. Hardcastle (Hrsg.), *Speech production and speech modelling* (Bd. 55). Dordrecht [u.a.], Niederlande: Kluwer [u.a.].

Marshall, I., Maran, N. J., Martin, S., Jan, M. A., Rimmington, J. E., Best, J. J., ... Douglas, N. J. (1993). Acoustic reflectometry for airway measurements in man: implementation and validation. *Physiological Measurement*, *14*, 157-169.

Maurer, D., Cook, N., Landis, T. & d'Heureuse, C. (1991). Are Measured Differences Between the Formants of Men, Women and Children Due to F0 Differences? *Journal of the International Phonetic Association*, *21* (2), 66-79.

Maurer, D., D'heureuse, C. & Landis, T. (2000). Formant Pattern Ambiguity of Vowel Sounds. *International Journal of Neuroscience*, *100* (1-4), 39-76.

Maurer, D., Hess, M. & Gross, M. (1996). High-speed imaging of vocal fold vibrations and larynx movements within vocalizations of different vowels. *The Annals of otology, rhinology & laryngology*, *105* (12), 975-981.

Maurer, D. & Klinkert, A. (1997). The spectral difference of different vowels: Toward a new acoustical concept. *The Journal of the Acoustical Society of America*, *101* (5), 3112. Zugriff auf `http://dx.doi.org/doi/10.1121/1.418889`

Maurer, D. & Landis, T. (1995). F0-Dependence, Number Alteration, and Non-Systematic Behaviour of the Formants in German Vowels. *International Journal of Neuroscience*, *83* (1-2), 25-44.

Maurer, D. & Landis, T. (1996). Intelligibility and spectral differences in high-pitched vowels. *Folia Phoniatrica et Logopaedica*, *48* (1), 1-10.

Maurer, D., Landis, T. & D'heureuse, C. (1991). Formant Movement and Formant Number Alteration With Rising F0 in Real Vocalizations of the German Vowels [u:], [o:] and [a:]. *International Journal of Neuroscience*, *57* (1-2), 25-38. Zugriff auf `http://informahealthcare.com/doi/abs/10.3109/00207459109150344`

Max, L. & Mueller, P. B. (1996). Speaking Fo and cepstral periodicity analysis of conversational speech in a 105-year-old woman: variability of aging effects. *Journal of Voice*, *10* (3), 245-251. Zugriff auf `http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=8865095`

Max, L., Wallace, M. E. & Vincent, I. (2003). Sensorimotor adaptation to auditory perturbations during speech: Acoustic and kinematic experiments. *Proceedings of the XV ICPHS Barcelona*.

McCaffrey, H. A. & Sussman, H. M. (1994). An Investigation of Vowel Organization in Speakers With Severe and Profound Hearing Loss. *Journal of Speech and Hearing Research*, *37* (4), 938-951. Zugriff auf `http://jslhr.asha.org/cgi/content/abstract/37/4/938`

McFarland, D., Baum, S. & Chabot, C. (1996). Speech compensation to structural modifications of the oral cavity. *The Journal of the Acoustical Society of America*, *100* (2), 1093-1104. Zugriff auf `http://dx.doi.org/10.1121/1.416286`

Melcon, M. C., Hoit, J. D. & Hixon, T. J. (1989). Age and laryngeal airway resistance during vowel production. *The Journal of speech and hearing disorders*, *54* (2), 282-286. Zugriff auf `http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=2709846`

Ménard, L., Perrier, P. & Savariaux, C. (2004). Exploring production perception relationships for 4 year old children: A study of compensation strategies to a lip tube perturbation. *The Journal of the Acoustical Society of America*, *115*, 2629.

Ménard, L., Schwartz, J.-L., Boë, L.-J. & Aubin, J. (2007). Articulatory-acoustic relationships during vocal tract growth for French vowels: Analysis of real data and simulations with an articulatory model. *Journal of Phonetics*, *35* (1), 1-19.

Ménard, L., Schwartz, J.-L., Boë, L.-J., Kandel, S. & Vallée, N. (2002). Auditory normalization of French vowels synthesized by an articulatory model simulating growth from birth to adulthood. *The Journal of the Acoustical Society of America*, *111* (4), 1892-1905. Zugriff auf `http://link.aip.org/link/?JAS/111/1892/1`

Mendel, L., Hamill, B., Crepeau, L. & Fallon, E. (1995). Speech intelligibility assessment in a environment. *The Journal of the Acoustical Society of America*, *97* (1), 628-636. Zugriff auf http://dx.doi.org/10.1121/1.412284

Mendes-Laureano, J., Sá, M. F. S., Ferriani, R. A., Reis, R. M., Aguiar-Ricz, L. N., Valera, F. C. P., ... Romão, G. S. (2006). Comparison of fundamental voice frequency between menopausal women and women at menacme. *Maturitas*, *55* (2), 195-199.

Meurer, E. M., Osório Wender, M. C., von Eye Corleta, H. & Capp, E. (2004). Phono-articulatory variations of women in reproductive age and postmenopausal. *Journal of Voice*, *18* (3), 369-374. Zugriff auf http://www.sciencedirect.com/science/article/B7585-4D5G579-K/2/5aa6b1118c092bdcc5cb916d9923a0f5

Miller, D. & Schutte, H. K. (1994). Toward a definition of male "head" register, passoggio, and "cover" in western operatic singing. *Folia Phoniatrica Logopaedica*, *46*, 157-170.

Miller, J. D. (1989). Auditory-perceptual interpretation of the vowel. *The Journal of the Acoustical Society of America*, *85* (5), 2114-2134.

Miller, R. L. (1953). Auditory tests with synthetic vowels. *Journal of the Acoustical Society of America*, *25* (114-121), 21.

Mooshammer, C. (2010). Acoustic and laryngographic measures of the laryngeal reflexes of linguistic prominence and vocal effort in German. *The Journal of the Acoustical Society of America*, *127* (2), 1047-1058. Zugriff auf http://dx.doi.org/10.1121/1.3277160

Mooshammer, C., Hoole, P., Alfonso, P. & Fuchs, S. (2001). Intrinsic pitch in German: A puzzle? *The Journal of the Acoustical Society of America*, *110*, 2761.

Moulines, E. & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication*, *9* (5-6), 453-467. Zugriff auf http://www.sciencedirect.com/science/article/B6V1C-48V21PK-GV/2/44631563bec26c612aa6c220164d71d1

Mullenix, J. W., Pisoni, D. B. & Martin, C. S. (1989). Some Effects of Talker Variability on Spoken Word Recognition. 1989. *Journal of the Acoustical Society of America*, *85* (1), 365-378.

Müller, V., Harrington, J., Kleber, F. & Reubold, U. (2011). Age-dependent differences in the neutralization of the intervocalic voicing contrast: Evidence from an apparent-time study on East Franconian. In *Proceedings of the 12th Annual Conference of the International Speech Communication Association (Interspeech2011)* (S. 633-636). Florenz, Italien.

Munhall, K. G., Lofqvist, A. & Kelso, J. A. S. (1994). Lip–larynx coordination in speech: Effects of mechanical perturbations to the lower lip. *The Journal of the Acoustical Society of America*, *95* (6), 3605-3616. Zugriff auf http://link.aip.org/link/?JAS/95/3605/1/http://dx.doi.org/10.1121/1.409929

Munhall, K. G., MacDonald, E. N., Byrne, S. K. & Johnsrude, I. (2009). Talkers alter vowel production in response to real-time formant perturbation even when instructed not to compensate. *The Journal of the Acoustical Society of America*, *125* (1), 384-390. Zugriff auf http://link.aip.org/link/?JAS/125/384/1/http://dx.doi.org/10.1121/1.3035829

Mwangi, S., Spiegl, W., Hönig, F., Haderlein, T., Maier, A. & Nöth, E. (2009). Effects of Vocal Aging on Fundamental Frequency and Formants. In Acoustical Society of the Netherlands & German Acoustical Society (Hrsg.), *Proceedings of the International Conference on Acoustics NAG/DAGA 2009* (S. 1761–1764). Rotterdam, Niederlande. Zugriff auf `http://www5.informatik.uni-erlangen.de/Forschung/Publikationen/2009/Mwangi09-EOV.pdf`

Mysak, E. D. & Hanley, T. D. (1958). Aging process in speech: Pitch and duration characteristics. *Journal of gerontology*, *13*, 309-313.

Nasir, S. M. & Ostry, D. J. (2006). Somatosensory Precision in Speech Production. *Current Biology*, *16* (19), 1918-1923. Zugriff auf `http://www.sciencedirect.com/science/article/pii/S096098220602001X`

Natke, U., Donath, T. M. & Kalveram, K. T. (2003). Control of voice fundamental frequency in speaking versus singing. *Journal of the Acoustical Society of America*, *113* (3), 1587-1593.

Nearey, T. M. (1978). *Phonetic feature systems for vowels.* Indiana University Linguistics Club.

Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *The Journal of the Acoustical Society of America*, *85*, 2088-2113.

Niebuhr, O. (2004). Intrinsic pitch in opening and closing diphthongs of German. *Proceedings of Speech Prosody 2004*, 733-736.

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of language and social psychology*, *18* (1), 62-85.

Nishio, M. & Niimi, S. (2008). Changes in fundamental frequency characteristics with aging. *Folia Phoniatrica et Logopaedica*, *60* (3), 120-127. Zugriff auf `http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=18305390`

Office of Public Sector Information - Information Policy Team. (2011). *The Queen's Christmas Broadcasts to the Commonwealth.* Zugriff auf `http://www.royal.gov.uk/ImagesandBroadcasts/TheQueensChristmasBroadcasts/AhistoryofChristmasBroadcasts.aspx`

Ogden, C. L., Fryar, C. D., Carroll, M. D. & Flegal, K. M. (2004). Mean body weight, height, and body mass index, United States 1960-2002. *Advanced Data from Vital and Health Statistics* (347), 1-17. Zugriff auf `http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=15544194`

Ohala, J. J. (1972). How is Pitch Lowered? *The Journal of the Acoustical Society of America*, *52* (1A), 124. Zugriff auf `http://scitation.aip.org/content/asa/journal/jasa/52/1A/10.1121/1.1981808`

Ohala, J. J. (1973). Explanations for the intrinsic pitch of vowels. *Monthly Internal Memorandum, Phonology Lab, University of California, Berkeley* (Januar), 9-26.

Ohala, J. J. & Eukel, B. W. (1987). Explaining the intrinsic pitch of vowels. In R. Channon & L. Shockey (Hrsg.), *In honour of Ilse Lehiste.* Dordrecht, Niederlande: Foris.

Pape, D. (2005). Is pitch perception and discrimination of vowels language-dependent and influenced by the vowels spectral properties? In *The Proceedings of the Ele-*

venth Meeting of the International Conference on Auditory Display, Limerick, Ireland (S. 340-343). Limerick, Irland. Zugriff auf `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.131.1277&rep=rep1&type=pdf`

Pape, D. & Mooshammer, C. (2004). Intrinsic pitch in German - Examining the whole fundamental frequency contour of the vowel. In *Proceedings of DAGA* (S. 897-898). Straßburg, Frankreich.

Pape, D. & Mooshammer, C. (2006a). Intrinsic F0 differences for German tense and lax vowels. In *Proceedings of the 7th International Seminar on Speech Production.* Ubatuba, Brasilien.

Pape, D. & Mooshammer, C. (2006b). Is Intrinsic Pitch language-dependent? Evidences from a Cross-Linguistic Vowel Pitch Experiment (with Additional Screening of the Listeners' DL for Music and Speech. In *Proceedings of Multilingual Speech and Language Processing (MULTILING-2006).* Stellenbosch, Südafrika.

Pape, D. & Mooshammer, C. (2008). Intrinsic pitch is not a universal phenomenon: Evidence from Romance languages. In *Proceedings of the 11th Labphon (Laboratory Phonology) Conference.* Wellington, New Zealand: University of Wellington.

Pape, D., Mooshammer, C., Fuchs, S. & Hoole, P. (2005). Intrinsic Pitch Differences Between German Vowels/i:/, /I/and /y:/ in a Crosslinguistic Perception Experiment. In *ISCA Workshop on Plasticity in Speech Perception.* London, Vereinigtes Königreich: University College London.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3. Aufl.). London [u.a.], Vereinigtes Königreich: Wadsworth Thomson Learning.

Pegoraro Krook, M. I. (1988). Speaking fundamental frequency characteristics of normal Swedish subjects obtained by glottal frequency analysis. *Folia phoniatrica*, *40* (2), 82-90. Zugriff auf `http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=3049278`

Pemberton, C., McCormack, P. & Russell, A. (1998). Have women's voices lowered across time? A cross sectional study of Australian women's voices. *Journal of Voice*, *12* (2), 208-213. Zugriff auf `http://www.sciencedirect.com/science?_ob=GatewayURL&_origin=ScienceSearch&_method=citationSearch&_piikey=S0892199798800404&_version=1&_returnURL=&md5=a5af7b376da232490046facb23db50ec`

Perkell, J. S. (2010). Movement goals and feedback and feedforward control mechanisms in speech production. *Journal of Neurolinguistics*, 1-26.

Perkell, J. S., Denny, M., Lane, H., Guenther, F., Matthies, M. L., Tiede, M., ... Burton, E. (2007). Effects of masking noise on vowel and sibilant contrasts in normal-hearing speakers and postlingually deafened cochlear implant users. *The Journal of the Acoustical Society of America*, *121* (1), 505-518. Zugriff auf `http://dx.doi.org/doi/10.1121/1.2384848`

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M. & Zandipour, M. (2004). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. *The Journal of the Acoustical Society of America*, *116* (4 Pt 1), 2338-2344. Zu-

griff auf http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=15532664

Perkell, J. S., Lane, H., Denny, M., Matthies, M. L., Tiede, M., Zandipour, M., ... Burton, E. (2007). Time course of speech changes in response to unanticipated short-term changes in hearing state. *The Journal of the Acoustical Society of America*, *121* (4), 2296-2311.

Perkell, J. S., Matthies, M. L., Svirsky, M. A. & Jordan, M. I. (1993). Trading relations between tongue-body raising and lip rounding in production of the vowel /u/: A pilot "motor equivalence" study. *The Journal of the Acoustical Society of America*, *93* (5), 2948-2961. Zugriff auf http://link.aip.org/link/?JAS/93/2948/1/http://dx.doi.org/10.1121/1.405814

Perkell, J. S., Matthies, M. L., Tiede, M., Lane, H., Zandipour, M., Marrone, N., ... Guenther, F. H. (2004). The Distinctness of Speakers' /s/–/sh/ Contrast Is Related to Their Auditory Discrimination and Use of an Articulatory Saturation Effect. *Journal of Speech, Language & Hearing Research*, *47* (6), 1259-1269. Zugriff auf http://jslhr.asha.org/cgi/content/abstract/47/6/1259

Peterson, G. E. (1952). The information-bearing elements of speech. *The Journal of the Acoustical Society of America*, *24* (6), 629-637. Zugriff auf http://dx.doi.org/10.1121/1.1906945

Peterson, G. E. (1961). Parameters of vowel quality. *Journal of Speech and Hearing Research*, *4* (1), 10.

Peterson, G. E. & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, *24* (2), 175-184.

Pörschmann, C. (2000). Influences of bone conduction and air conduction on the sound of one's own voice. *Acta Acustica united with Acustica*, *86* (6), 1038-1045.

Potter, R. & Steinberg, J. (1950). Toward the Specification of Speech. *The Journal of the Acoustical Society of America*, *22* (6), 807.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (1992). *Numerical recipes in C: The art of scientific computing* (2. Aufl.). Cambridge [u.a.]: Cambridge University Press.

Purcell, D. W., Kunov, H. & Cleghorn, W. (2003). Estimating bone conduction transfer functions using otoacoustic emissions. *The Journal of the Acoustical Society of America*, *114*, 907-918.

Purcell, D. W. & Munhall, K. G. (2006a). Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation. *The Journal of the Acoustical Society of America*, *120* (2), 966-977. Zugriff auf http://link.aip.org/link/?JAS/120/966/1

Purcell, D. W. & Munhall, K. G. (2006b). Compensation following real-time manipulation of formants in isolated vowels. *The Journal of the Acoustical Society of America*, *119* (4), 2288-2297. Zugriff auf http://link.aip.org/link/?JAS/119/2288/1

R Development Core Team. (2010). *R: A Language and Environment for Statistical Computing.* Wien, Österreich. Zugriff auf http://www.R-project.org

Ramig, L. A. & Ringel, R. L. (1983). Effects of physiological aging on selected acoustic

characteristics of voice. *Journal of Speech and Hearing Research*, *26* (1), 22-30. Zugriff auf http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=6865377

Rastatter, M. P. & Jacques, R. D. (1990). Formant frequency structure of the aging male and female vocal tract. *Folia phoniatrica*, *42* (6), 312-319.

Rastatter, M. P., McGuire, R. A., Kalinowski, J. & Stuart, A. (1997). Formant frequency characteristics of elderly speakers in contextual speech. *Folia Phoniatrica et Logopaedica*, *49* (1), 1-8. Zugriff auf http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=9097490

Reinholt Petersen, N. (1978). Intrinsic Fundamental Frequency of Danish Vowels. *Journal of Phonetics*, *6* (3), 177-189.

Reinholt Petersen, N. (1980). Coarticulation of inherent fundamental frequency levels between syllables. *Annual Reports of the Institute of Phonetics, University of Copenhagen*, *14*, 317-354.

Reinholt Petersen, N. (1986). Perceptual Compensation for Segmentally Conditioned Fundamental Frequency Perturbation. *Phonetica*, *43* (1-3), 31-42.

Reubold, U., Harrington, J. & Kleber, F. (2010). Vocal aging effects on F0 and the first formant: A longitudinal analysis in adult speakers. *Speech Communication*, *52* (7-8), 638-651. Zugriff auf http://www.sciencedirect.com/science/article/B6V1C-4YH4PTG-1/2/5832d1873d76a6ce47580ad9bd5c1c46

Rhodes, R. (2011). *Changes in the voice across the adult lifespan.* Zugriff auf http://www.kfs.oeaw.ac.at/publications/iafpa_abstracts/nr31_rrhodes_revised.pdf

Richards, D. W. (1965). Pulmonary changes due to aging. In R. Fenn & H. Rahn (Hrsg.), *Handbook of physiology. A critical comprehensive presentation of physiologic knowledge and concepts.* (Bd. 3, S. 1525-1529). Washington, D.C., USA: American Physiological Society.

Riordan, C. J. (1977). Control of vocal-tract length in speech. *The Journal of the Acoustical Society of America*, *62* (4), 998-1002. Zugriff auf http://link.aip.org/link/?JAS/62/998/1/http://dx.doi.org/10.1121/1.381595

Rostolland, D. (1982). Phonetic structure of shouted voice. *Acustica*, *51*, 80-89.

Russell, A., Penny, L. & Pemberton, C. (1995). Speaking Fundamental Frequency Changes Over Time in Women: A Longitudinal Study. *Journal of Speech and Hearing Research*, *38* (1), 101-109. Zugriff auf http://jslhr.asha.org/cgi/content/abstract/38/1/101

Ryalls, J. H. & Lieberman, P. (1982). Fundamental frequency and vowel perception. *The Journal of the Acoustical Society of America*, *72* (5), 1631-1634. Zugriff auf http://link.aip.org/link/?JAS/72/1631/1/http://dx.doi.org/10.1121/1.388499

Sankoff, G. & Blondeau, H. (2007). Language change across the lifespan: /r/ in Montreal French. *Language*, *83* (3), 560-588.

Sato, K. & Hirano, M. (1997). Age-related changes of elastic fibers in the superficial layer of the lamina propria of vocal folds. *The Annals of otology, rhinology, and laryngology*, *106* (1), 44-48. Zugriff auf http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=9006361

Savariaux, C., Perrier, P. & Ohteru, S. (1995). Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube: A study of the control space in speech production. *The Journal of the Acoustical Society of America*, *98*, 2428.

Savariaux, C., Perrier, P., Orliaguet, J.-P. & Schwartz, J.-L. (1999). Compensation strategies for the perturbation of French [u] using a lip tube. II. Perceptual analysis. *The Journal of the Acoustical Society of America*, *106*, 381-393.

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

Schiel, F. (1999). Automatic phonetic transcription of non-prompted speech. In *Proceedings of the XIVth International Congress of Phonetic Sciences (ICPhS).* San Francisco.

Schiel, F. (2004). MAuS goes iterative. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC).* Lissabon, Portugal.

Schneider, B., van Trotsenburg, M., Hanke, G., Bigenzahn, W. & Huber, J. (2004). Voice impairment and menopause. *Menopause*, *11* (2), 151-158.

Schulman, R. (1985a). Articulatory Targeting and perceptual constancy of loud speech. In *PERILUS 4* (S. 86-91). Stockholm, Schweden: Institute of Linguistics, Universität Stockholm.

Schulman, R. (1985b). Dynamic and perceptual constraints of loud speech. *The Journal of the Acoustical Society of America*, *78*, S37.

Schulman, R. (1989). Articulatory dynamics of loud and normal speech. *The Journal of the Acoustical Society of America*, *85* (1), 295-312. Zugriff auf `http://link.aip.org/link/?JAS/85/295/1`

Scukanec, G. P., Petrosino, L. & Squibb, K. (1991). Formant frequency characteristics of children, young adult, and aged female speakers. *Perceptual and motor skills*, *73* (1), 203-208.

Segre, R. (1971). Senescence of the voice. *Eye, ear, nose & throat monthly*, *50* (6), 223-227. Zugriff auf `http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=5089667`

Shiller, D. M., Sato, M., Gracco, V. L. & Baum, S. R. (2009). Perceptual recalibration of speech sounds following speech motor learning. *The Journal of the Acoustical Society of America*, *125* (2), 1103-1113.

Shipp, T. (1975). Vertical laryngeal position during continuous and discrete vocal frequency change. *Journal of Speech and Hearing Research*, *18* (4), 707-718. Zugriff auf `http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=1207101`

Shue, Y.-L. (2011). *VoiceSauce: a program for voice analysis.* Zugriff auf `http://www.ee.ucla.edu/~spapl/voicesauce/`

Shue, Y.-L., Chen, G. & Alwan, A. (2010). On the Interdependencies between Voice Quality, Glottal Gaps, and Voice-Source related Acoustic Measures. In *Proceedings of INTERSPEECH 2010* (S. 34-37). Makuhari, Japan.

Siegel, G. M., Pick, H. L., Olsen, M. G. & Sawin, L. (1976). Auditory feedback on the regulation of vocal intensity of preschool children. *Developmental Psychology*, *12* (3), 255.

Simpson, A. P. (2001). Dynamic consequences of differences in male and female vocal tract dimensions. *The Journal of the Acoustical Society of America*, *109*, 2153.

Slawson, A. W. (1968). Vowel quality and musical timbre as functions of spectrum envelope and fundamental frequency. *The Journal of the Acoustical Society of America*, *43* (1), 87-101. Zugriff auf `http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=5636408`

Smith, A. (2010). Development of neural control of orofacial movements for speech. In W. J. Hardcastle, J. Laver & F. Gibbon (Hrsg.), *The Handbook of Phonetic Sciences.* Oxford, Vereinigtes Königreich: Wiley-Blackwell.

Sorensen, D. & Horii, Y. (1982). Cigarette smoking and voice fundamental frequency. *Journal of Communication Disorders*, *15* (2), 135-144.

Srivastava, S., Gupta, M. R. & Frigyik, B. A. (2007). Bayesian quadratic discriminant analysis. *Journal of Machine Learning Research*, *8*, 1287-1314.

Stenfelt, S. & Goode, R. L. (2005). Bone-Conducted Sound: Physiological and Clinical Aspects. *Otology & Neurotology*, *26* (6), 1245-1261. Zugriff auf `http://journals.lww.com/otology-neurotology/Fulltext/2005/11000/Bone_Conducted_Sound__Physiological_and_Clinical.31.aspx`

Stevens, K. N. (1971). Airflow and turbulence noise for fricative and stop consonants: static considerations. *The Journal of the Acoustical Society of America*, *50*, 1180-1192.

Stevens, K. N. & Keyser, S. J. (2010). Quantal theory, enhancement and overlap: Phonetic Bases of Distinctive Features. *Journal of Phonetics*, *38* (1), 10-19. Zugriff auf `http://www.sciencedirect.com/science/article/pii/S0095447008000600`

Stoicheff, M. L. (1981). Speaking Fundamental Frequency Characteristics of Nonsmoking Female Adults. *Journal of Speech and Hearing Research*, *24* (3), 437-441. Zugriff auf `http://jslhr.asha.org/cgi/content/abstract/24/3/437`

Stoll, G. (1984). Pitch of vowels: Experimental and theoretical investigation of its dependence on vowel quality. *Speech Communication*, *3* (2), 137-147. Zugriff auf `http://www.sciencedirect.com/science/article/B6V1C-48TN37G-3/2/2d7a1326afd5814421ccd27387ab3c31`

Sundberg, J. (1975). Formant technique in a professional female singer. *Acoustica*, *32*, 89-96.

Sundberg, J. (1977). The acoustics of the singing voice. *Scientific American*, *236* (3), 82-91.

Sundberg, J. (1987). *The science of the singing voice* (Bd. 1). Dekalb, Illinois: Northern Illinois University Press.

Sundberg, J. & Nordström, P.-E. (1976). Raised and lowered larynx - the effect on vowel formant frequencies. *KTH - Quarterly Progress and Status Report*, 35-39. Zugriff auf `http://www.speech.kth.se/prod/publications/files/qpsr/1976/1976_17_2-3_035-039.pdf`

Sussman, H. M. (1986). A neuronal model of vowel normalization and representation. *Brain and Language*, *28* (1), 12-23.

Syrdal, A. K. & Gopal, H. S. (1983). Perceived critical distances between F1 - F0, F2 - F1, F3 - F2. *The Journal of the Acoustical Society of America*, *74* (S1),

S88-S89. Zugriff auf http://link.aip.org/link/?JAS/74/S88/5/http://dx.doi.org/10.1121/1.2021201

Syrdal, A. K. & Gopal, H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *The Journal of the Acoustical Society of America*, *79* (4), 1086-1100. Zugriff auf http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=3700864

Syrdal, A. K. & Steele, S. A. (1985). Vowel F1 as a function of speaker fundamental frequency. *The Journal of the Acoustical Society of America*, *78* (S1), S56. Zugriff auf http://link.aip.org/link/?JAS/78/S56/3

Tabain, M. & Perrier, P. (2007). An articulatory and acoustic study of /u/ in preboundary position in French: The interaction of compensatory articulation, neutralization avoidance and featural enhancement. *Journal of Phonetics*, *35* (2), 135-161. Zugriff auf http://www.sciencedirect.com/science/article/B6WKT-4M2WNX2-1/2/6140629692f75c59994665a2bc387b8e

Thorsen, N. (1984). Variability and invariance in Danish stress group patterns. *Phonetica*, *41* (2), 88-102.

Titze, I. R. (2004). A theoretical study of f0-f1 interaction with application to resonant speaking and singing voice. *Journal of Voice*, *18* (3), 292-298. Zugriff auf http://www.sciencedirect.com/science/article/B7585-4D5G579-8/2/d83e0097b8e5eca3881e40ac124bec05

Titze, I. R. & Sundberg, J. (1992). Vocal intensity in speakers and singers. *The Journal of the Acoustical Society of America*, *91*, 2936.

Torre III, P. & Barlow, J. A. (2009). Age-related changes in acoustic characteristics of adult speech. *Journal of Communication Disorders*, *42* (5), 324-333. Zugriff auf http://www.sciencedirect.com/science/article/B6T85-4VYXMN7-1/2/adb3312a6e7987663b4019647ee7467a

Traunmüller, H. (1981). Perceptual dimension of openness in vowels. *The Journal of the Acoustical Society of America*, *69* (5), 1465-1475.

Traunmüller, H. (1983). *On Vowels: Perception of Spectral Features, Related Aspects of Production and Sociophonetic Dimensions: Univ., Diss.–Stockholm, 1983.* University of Stockholm.

Traunmüller, H. (1984). Articulatory and perceptual factors controlling the age- and sex-conditioned variability in formant frequencies of vowels. *Speech Communication*, *3* (1), 49-61. Zugriff auf http://www.sciencedirect.com/science/article/B6V1C-4998RVX-8/2/c09eecb2f63ad2c9df1bea0aae9f52dd

Traunmüller, H. (1985). The role of the fundamental and the higher formants in the perception of speaker size, vocal effort, and vowel openness. In *PERILUS 4* (S. 92-102). Stockholm, Schweden: Institute of Linguistics, Universität Stockholm.

Traunmüller, H. (1988). Paralinguistic variation and invariance in the characteristic frequencies of vowels. *Phonetica*, *45* (1), 1-29.

Traunmüller, H. (1990a). Analytical expressions for the tonotopic sensory scale. *The Journal of the Acoustical Society of America*, *88* (1), 97-100.

Traunmüller, H. (1990b). A note on hidden factors in vowel perception experiments. *The Journal of the Acoustical Society of America*, *88* (4), 2015-2019.

Traunmüller, H. (1991a). The context sensitivity of the perceptual interaction between F0 and F1. In *Proceedings of the 12th International Congress of Phonetic Sciences Aix-en-Provence* (S. 62-65). Aix-en-Provence, Frankreich. Zugriff auf `http://www.ling.su.se/staff/hartmut/Aix_91.htm`

Traunmüller, H. (1991b). Function and limits of the F1: F0 covariation in speech. In *PERILUS XIV* (S. 125-129). Stockholm, Schweden: Stockholm University. Zugriff auf `http://www2.ling.su.se/fon/perilus/perilus14_1991.pdf#page=155`

Traunmüller, H. (2005). *Speech considered as modulated voice.* Zugriff am 19.09.2011 auf `http://www2.ling.su.se/staff/hartmut/speech_considered.pdf`

Traunmüller, H. & Eriksson, A. (1995a). *The frequency range of the voice fundamental in the speech of male and female adults.* Zugriff auf `http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-10290`

Traunmüller, H. & Eriksson, A. (1995b). The perceptual evaluation of F0 excursions in speech as evidenced in liveliness estimations. *The Journal of the Acoustical Society of America*, *97* (3), 1905-1915. Zugriff auf `http://dx.doi.org/10.1121/1.412942`

Traunmüller, H. & Eriksson, A. (2000). Acoustic effects of variation in vocal effort by men, women, and children. *The Journal of the Acoustical Society of America*, *107* (6), 3438-3451.

Traunmüller, H. & Lacerda, F. (1987). Perceptual relativity in identification of two-formant vowels. *Speech Communication*, *6* (2), 143-157.

Turk, A. & Sawusch, J. R. (1996). The processing of duration and intensity cues to prominence. *The Journal of the Acoustical Society of America*, *99* (6), 3782-3890.

Turner, R. E. & Patterson, R. D. (2003). An analysis of the size information in classical formant data: Peterson and Barney (1952) revisited. *Journal of the Acoustical Society of Japan*, *33* (9), 585-589.

Turner, R. E., Walters, T. C., Monaghan, J. J. M. & Patterson, R. D. (2009). A statistical, formant-pattern model for segregating vowel type and vocal-tract length in developmental formant data. *The Journal of the Acoustical Society of America*, *125* (4), 2374-2386. Zugriff auf `http://scitation.aip.org/content/asa/journal/jasa/125/4/10.1121/1.3079772`

Tye-Murray, N. (1987). Effects of Vowel Context on the Articulatory Closure Postures of Deaf Speakers. *Journal of Speech and Hearing Research*, *30* (1), 99-104. Zugriff auf `http://jslhr.asha.org/cgi/content/abstract/30/1/99`

Vallabha, G. K. & Tuller, B. (2002). Systematic errors in the formant analysis of steady-state vowels. *Speech Communication*, *38* (1-2), 141-160. Zugriff auf `http://www.sciencedirect.com/science/article/pii/S0167639301000498`

van Dommelen, W. A. (1990). Acoustic parameters in human speaker recognition. *Language and Speech*, *33* (3), 259-272. Zugriff auf `http://las.sagepub.com/content/33/3/259.short`

van Dommelen, W. A. (1993). Speaker height and weight identification: a reevaluation of some old data. *Journal of Phonetics*, *21*, 337-341. Zugriff auf `http://psycnet.apa`

.org/index.cfm?fa=search.displayRecord&UID=1994-00390-001

Van Hoof, S. & Verhoeven, J. (2011). Intrinsic vowel F0, the size of vowel inventories and second language acquisition. *Journal of Phonetics*, *39* (2), 168-177. Zugriff auf http://www.sciencedirect.com/science/article/pii/S0095447011000234

Venables, W. N. & Ripley, B. D. (2002). *Modern Applied Statistics with S* (4. Aufl.). New York: Springer. Zugriff auf http://www.stats.ox.ac.uk/pub/MASS4

Verdonck-De Leeuw, I. M. & Mahieu, H. F. (2004). Vocal aging and the impact on daily life: a longitudinal study. *Journal of Voice*, *18* (2), 193-202. Zugriff auf http://linkinghub.elsevier.com/retrieve/pii/S089219970300153X?showall=true

Verhoeven, J. & Van Hoof, S. (2007). Intrinsic vowel pitch in Dutch and Arabic. In *Proceedings of the XVIth International Conference on Phonetic Sciences* (S. 1785-1788). Saarbrücken.

Vilkman, E., Aaltonen, O., Raimo, I., Arajärvi, P. & Oksanen, H. (1989). Articulatory hyoid-laryngeal changes vs. cricothyroid muscle activity in the control of intrinsic F0 of vowels. *Journal of Phonetics*, *17*, 193-203.

Villacorta, V. M. (2006). *Sensorimotor adaptations to perturbations of vowel acoustics and its relation to perception* (Unveröffentlichte Dissertation). MIT.

Villacorta, V. M., Perkell, J. S. & Guenther, F. H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *The Journal of the Acoustical Society of America*, *122* (4), 2306-2319. Zugriff auf http://link.aip.org/link/?JAS/122/2306/1/http://dx.doi.org/10.1121/1.2773966

von Helmholtz, H. (1865). *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik*. Braunschweig: Vieweg.

Watson, P. J. & Munson, B. (2007). A Comparison of Vowel Acoustics Between Older and Younger Adults. In *Proceedings of the XVIth International Congress of Phonetic Sciences* (S. 561-564). Saarbrücken.

Wempe, T. & Boersma, P. (2003). The interactive designs of an f0-related spectral analyser. In *Proceedings of the Institute of Phonetic Sciences, Amsterdam* (Bd. 25, S. 163-170). Amsterdam, Niederlande.

Whalen, D. H., Gick, B., Kumada, M. & Honda, K. (1999). Cricothyroid activity in high and low vowels: exploring the automaticity of intrinsic F0. *Journal of Phonetics*, *27* (2), 125-142.

Whalen, D. H. & Levitt, A. G. (1995). The universality of intrinsic F0 of vowels. *Journal of Phonetics*, *23* (3), 349-366.

Whiteside, S. P. (2001). Sex-specific fundamental and formant frequency patterns in a cross-sectional study. *The Journal of the Acoustical Society of America*, *110* (1), 464-478.

Wilder, C. (1978). Vocal aging. In *Transcriptions on the seventh symposion care of the professional voice.* New York, USA: The Voice Foundation.

Wind, J. (1970). *On the phylogeny and the ontogeny of the human larynx: A morphological and functional study*. Groningen: Wolters-Noordhoff.

Wolpert, D. M., Ghahramani, Z. & Flanagan, J. (2001). Perspectives and problems in motor learning. *Trends in Cognitive Sciences*, *5* (11), 487-494. Zugriff auf http://

www.sciencedirect.com/science/article/pii/S1364661300017733

Wood, S. (1986). The acoustical significance of tongue, lip, and larynx maneuvers in rounded palatal vowels. *The Journal of the Acoustical Society of America*, *80* (2), 391-401. Zugriff auf http://link.aip.org/link/?JAS/80/391/1

Xia, K. & Espy-Wilson, C. (2000). A new strategy of formant tracking based on dynamic programming. In *Proceedings of the 6th International Conference on Spoken Language Processing.* Peking, China.

Xu, Y., Larson, C., Bauer, J. & Hain, T. (2004). Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences. *The Journal of the Acoustical Society of America*, *116* (2), 1168-1178.

Xue, S. A. & Hao, G. J. (2003). Changes in the Human Vocal Tract Due to Aging and the Acoustic Correlates of Speech Production: A Pilot Study. *Journal of Speech, Language & Hearing Research*, *46* (3), 689-701. Zugriff auf http://jslhr.pubs.asha.org/article.aspx?articleid=1781231

Xue, S. A., Jiang, J., Lin, E., Glassenberg, R. & Mueller, P. B. (1999). Age-related changes in human vocal tract configurations and the effects on speakers' vowel formant frequencies: a pilot study. *Logopedics Phoniatrics Vocology*, *24* (3), 132–137.

Yates, A. J. (1963). Delayed auditory feedback. *Psychological Bulletin*, *60* (3), 213-232.

Zemlin, W. R. (1998). *Speech and hearing science: Anatomy and physiology* (4. Aufl.). Boston: Allyn and Bacon.

Zharkova, N., Hewlett, N. & Hardcastle, W. J. (2011). Coarticulation as an Indicator of Speech Motor Control Development in Children: An Ultrasound Study. *Motor Control*, *15* (1), 118-140.

Zwicker, E. (1982). *Psychoakustik*. Berlin: Springer.