

Vorabdruck von: Siebenhaar, Beat (2015): "Quantitative Ansätze zu einer Sprachgeographie der schweizerdeutschen Prosodie". In: Kehrein, Roland, Alfred Lameli und Stefan Rabanus (Hg.): Regionale Variation des Deutschen – Projekte und Perspektiven. Berlin/Boston: De Gruyter: 195 – 217.

Beat Siebenhaar

## QUANTITATIVE ANSÄTZE ZU EINER SPRACHEGEOGRAPHIE DER SCHWEIZERDEUTSCHEN PROSODIE

### 1 Gegenstand

Die meisten innersprachlichen Ebenen der schweizerdeutschen Mundarten sind gut bis sehr gut untersucht, sowohl im Hinblick auf die Einzeldialekte als auch im sprachgeographischen Vergleich. Was zu Beginn des 21. Jahrhunderts fehlt, ist neben der syntaktischen Ebene, die z. Z. mit dem Syntaxatlas der deutschen Schweiz erfasst wird, eine Beschreibung der Prosodie. In diese Lücke ist das 2008 abgeschlossene Projekt eingedrungen. Eine umfassende oder auch eine geographisch breit abgestützte Beschreibung konnte damit aber nicht gegeben werden; das war auch nicht geplant, denn viele theoretische und methodische Problemfelder sind bei Weitem noch nicht gelöst. Erste Ergebnisse zeigen aber mögliche Ansätze zu einer Erforschung auf, die Methodendiskussion ist also Teil der Ergebnisse. Mit der Erhebung von Daten in vier – ursprünglich geplant waren sieben – Regionen wurden Grundmuster einer prosodischen Variation der dialektalen Spontansprache in der Schweiz erfasst, wobei situative und soziale Faktoren möglichst konstant gehalten wurden. Aus der Literatur konnte davon ausgegangen werden, dass in den alemannischen Mundarten der Schweiz, welche immer noch die Alltagssprache der Kommunikation unter Einheimischen darstellen, nicht grundsätzlich bzw. typologisch

unterschiedliche prosodische Muster vorliegen. Allerdings wurden besonders für die Walliser Dialekte, die sich segmental deutlich von den andern Deutschschweizer Dialekten unterscheiden, schon früh auch prosodische Besonderheiten genannt. Diese wurden wie anderenorts für prosodische Auffälligkeiten ‚Singen‘ genannt (Wipf 1910; generell zum ‚Singen‘ als Begriff der Prosodie: Gilles & Schrambke 2000, Siebenhaar 2012, Zimmermann 1998). Von phonologischen oder typologischen Unterschieden innerhalb der Mundarten, wie sie Fitzpatrick-Cole (1999) für die Differenzierung des Berndeutschen vom bundesdeutschen Standard festgestellt hatte, war also mit Ausnahme der Walliser Dialekte nicht auszugehen. Vielmehr war zu erwarten, dass die Variation eher auf phonetischer Ebene anzusetzen war: eine leicht unterschiedliche Position des Peaks des Akzents in der Silbe oder ein unterschiedliches Verhältnis von Lautdauern. Dieser Vorannahme entsprechend liegt der Fokus auf einer kleinteiligen phonetischen Analyse. Da diese Unterschiede zwar den holistischen Eindruck der Prosodie prägen, aber selbst durch eine große Variation ausgezeichnet sind, wird ein quantitativer Ansatz gewählt.

Aus Aufnahmen von 10 freien Gesprächen pro Ort werden die prosodisch am wichtigsten beurteilten Parameter – Zeit und Grundfrequenz – analysiert. Für jeden Erhebungsort und jede Gewährsperson werden Modelle für die Phrasierung, das Timing und die Intonation erstellt, geplante Bezüge der Teilmodelle sind aus Zeitgründen nur rudimentär ausgefallen. Der Vergleich der Orte ergibt einen Einblick in die regionale Variation der Prosodie. Der grundsätzlich quantitative Zugang erlaubt auch Modelle zu entwickeln, welche zeigen, wie stark die Prosodie durch den Faktor Raum bestimmt ist oder ob andere Faktoren (z. B. individueller Stil, Register, Geschlecht) bedeutsamer sind. Im Projektverlauf sind zusätzlich Analysen zur Perzeption

hinzugekommen, die zeigen, inwiefern die festgestellten Unterschiede von kompetenten Laien überhaupt wahrgenommen und diskriminiert werden.

### 1.1 Organisationsrahmen

Das am Institut für Sprachwissenschaft der Universität Bern angesiedelte Projekt ist von 2005 bis 2008 vom Schweizerischen Nationalfonds finanziert worden. Projektleiter war Iwar Werlen. Adrian Leemann und Beat Siebenhaar waren Projektmitarbeiter. Als Hilfskräfte haben Kathrin Keller Flückiger, Franziska Leuenberger, Jürg Bigler, Heidy Suter und Manuela Guntern Aufnahmen gemacht, transkribiert und gelabelt.

Das Projekt gründet auf einem Vorgängerprojekt, das von 2002–2005 an der Universität Lausanne von Eric Keller und an der Universität Bern von Iwar Werlen geleitet wurde. Ziel dieses Vorgängerprojekts (Mitarbeiter\_innen: Martin Forst, Katrin Häsler, Ingrid Hove, Beat Siebenhaar) war es, eine Sprachsynthese für zwei Deutschschweizer Mundarten zu erstellen. Eine vollständige Synthese konnte nicht erstellt werden, was an der Signalgeneration gescheitert ist. Die prosodische Modellierung für drei Sprecher von zwei Mundarten kann aber als gelungen bezeichnet werden und bildet die methodische Grundlage für das hier vorgestellte Projekt (Häsler, Hove & Siebenhaar 2005). Die Modellierung der drei Sprecher hat deutlich gemacht, dass individuelle Unterschiede die dialektalen Unterschiede so stark überlagern, dass nicht zu entscheiden war, welches dialektale und welches idiolektale Besonderheiten waren. Auf dieser Erkenntnis aufbauend verbreiterte das hier dargestellte Projekt die Datenbasis, insofern als mehrere Personen pro Ort aufgenommen wurden. Die Prinzipien der Modellierung sind

mit leichten, im Bereich der Intonation mit deutlicheren Modifikationen beibehalten worden.

## 1.2 Korpus und Methode

Im Zentrum der Untersuchung steht die Prosodie der Alltagssprache, also nicht archaische Mundartformen, wobei hier anzumerken ist, dass die Dialekte in der Schweiz immer noch die Alltagssprache darstellen und sich in phonetischer und morphologischer Hinsicht im Wesentlichen noch mit den Daten aus der Mitte des 20. Jahrhunderts decken (Christen 1998). Datenbasis für das hier beschriebene Projekt sind demnach spontansprachliche Interviews, die mit Schülern gymnasialer Abschlussklassen geführt wurden, womit die soziale und situative Variation minimiert wurde. Themen der Interviews waren der schulische Alltag, Zukunftspläne und Hobbies, zusätzlich wurden Bildbeschreibungen und eine Nacherzählung aufgenommen, die jedoch nicht analysiert wurden. Der allgemeinen Sprachsituation der Schweiz entsprechend wurden alle Interviews selbstverständlich im Dialekt geführt. Mit Ausnahme der Walliser Daten, für welche teilweise ein Subsidiärdialekt für die Sprache mit Außenstehenden angesetzt wird (Schnidrig 1986, Werlen u. a. 2002), ist nicht von allzu starken Akkommodationserscheinungen auszugehen, da die Dialekte, die alle dem Hoch- und Höchstalemannischen zugeordnet werden, untereinander gut verständlich sind. Dieses Verständnis wird durch den selbstverständlichen Gebrauch im Alltag, aber auch in den Massenmedien gefördert.

Für die statistische Analyse von Lautdauern und Akzenten wurden die Aufnahmen auf Lautebene<sup>1</sup> segmentiert. Damit ist jeder einzelne Laut in seiner Dauer bestimmt, die Lautsegmente wurden mit weiteren Informationen annotiert: Aus dem Vorgängerprojekt und aus der Literatur zur prosodischen Modellierung für die Sprachsynthese sind weitere phonetische und linguistische Parameter bekannt, welche für die Prosodie (mit-)bestimmend sind. Das sind insbesondere die grammatische Kategorie des Wortes (lexikalisch vs. grammatisch), die Position des Lautes in der Silbe (Onset, Nukleus, Coda), die vorangehenden und nachfolgenden Laute, der lexikalische Akzent der Silbe (*stressed*, *non-stressed*, Schwa-Silbe) und enger Fokus (ja vs. nein), Phrasengrenzen (iP, IP) und Phrasentyp (abschließend, weiterführend, fragend, Unterbruch). Für die Intonationsanalyse wurden weitere Parameter eingefügt, die jedoch erst nach einer ersten Analyse rekurrierend angesetzt werden konnten, wie die Länge der vorhergehenden Phrase und die Höhe ihrer Phrasenkomponente. Zudem werden für die Intonationsanalyse auch die Emotionalität berücksichtigt und pro Ort drei Klassen von Sprechern auf der Basis ihrer Sprechgeschwindigkeit erfasst.

Für die Berechnung von Unterschieden im temporalen Bereich können diese Parameter und direkt davon abgeleitete Variablen in ANOVAs eingesetzt

<sup>1</sup> Der etwas vage Begriff *Laut* ist an dieser Stelle gewählt, weil weder eine rein phonetische noch eine phonologische Annotation vorliegt. Der Begriff *Phon* ist nicht stimmig, weil auch eine phonetische Segmentation nicht ohne phonologische Information auskommt. Bei Plosiven sind zudem jeweils Okklusions- und Lösungsphase getrennt annotiert worden, wodurch die Einheit in diesem Fall kleiner als das Phon ist. Der intern verwendete Begriff *Segment* ist andererseits nichtssagend, weil auch Phrasen Segmente sind. Ausgangspunkt der Segmentation war jeweils eine phonologische Analyse, wobei auch Allophone markiert wurden, weshalb auch *Phonem* nicht verwendet werden kann. Ebenfalls sind Reduktionsformen, wie sie beim spontanen Sprechen immer vorkommen, jeweils in ihrer Vollform und der Reduktion festgehalten worden, so dass die notierte phonetische Einheit insgesamt immer auf die phonologische Grundlage zurückgeführt werden konnte (siehe zur problematischen Distinktion phonetisch-phonologischer Annotation insb. Siebenhaar / Leemann 2012: 26–29).

werden, welche als Standardverfahren der Variationslinguistik Hinweise auf messbare Unterschiede nachweisen. Die Dauerwerte werden aber nicht als Rohdaten verwendet, sondern mit dem Logarithmus zur Basis 10; die linkssteile Verteilung der Rohdaten kann so auf eine Gauß'sche Normalverteilung skaliert werden, welche Voraussetzung für verschiedene Analyseverfahren ist. Zusätzlich werden Lineare Modelle erstellt. Diese Modelle können v. a. die Interaktion zwischen verschiedenen Variablen erfassen und haben damit auch Erklärungspotenzial.

Für die Erfassung der Intonation mussten die Daten weiter aufgearbeitet werden, da die Grundfrequenz selbst (noch) nicht direkt für die Analyse verwendet werden kann. Die komplexen Intonationsverläufe müssen so vereinfacht werden, dass die Analyse auf eine überschaubare Anzahl von Variablen zurückgreifen kann, welche den  $F_0$ -Verlauf möglichst gut abbilden. Dafür wird im vorliegenden Projekt auf ein Command-Response-Modell zurückgegriffen, das von einem Team um Hiroya Fujisaki seit den 1970er-Jahren entwickelt worden ist (Fujisaki & Hirose 1982). Theoretischer Hintergrund des Modells ist die Annahme, dass Intonation in Form von globalen und lokalen  $F_0$ -Bewegungen hierarchisch strukturiert ist und sich der komplexe  $F_0$ -Verlauf einer Äußerung in Parametern wie Phrasenkommandos für den globalen Verlauf und Akzentkommandos für die lokalen Bewegungen fassen lässt, wie das in der Abbildung 1 dargestellt ist. Die beiden Kontrollmechanismen, die auf der linken Seite der Abbildung dargestellt sind, werden addiert und ergeben so die gesamte Kontur. Für die detaillierte Beschreibung des Modells ist hier auf Leemann (2012: 41–67) verwiesen

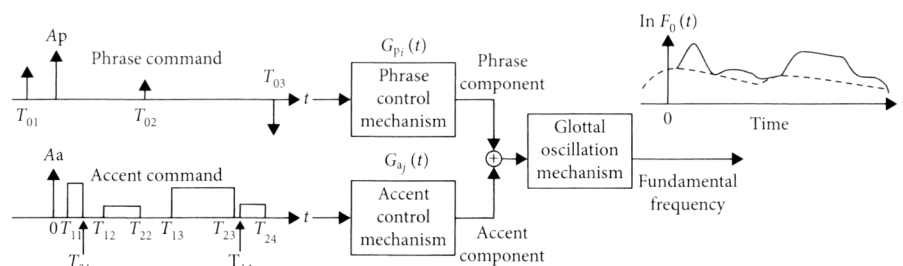


Abbildung 1: Blockdiagramm des Fujisaki-Intonationsmodell (Leemann 2012: 43 nach Fujisaki & Hirose 1984: 235)

Mittels eines von Mixdorff (2009) entwickelten (Labor-)Programms werden die Intonationsverläufe jeder einzelnen Phrase als Copysynthese abgebildet und in einer mathematischen Formel dargestellt. Die damit errechneten Parameter, welche den Intonationsverlauf nachbilden, werden in einem weiteren Schritt mit den linguistischen, paralinguistischen und außersprachlichen Faktoren korreliert. Für die globalen Phrasenkommandos sind das Zeitpunkt und Höhe des Phrasenkommandos, für lokale Akzentkommandos werden Beginn und Ende sowie die Höhe des Kommandos errechnet. Hinzu kommt die sprecherspezifische Basis der Grundfrequenz. Somit kann jeder Punkt einer Intonationskurve mit dem festen Parameter Grundfrequenz und fünf Variablen bestimmt werden. Dieses Verfahren der Analyse durch Synthese hat den Vorteil, dass die Abstraktion der Intonation sehr nah an der phonetischen Realität des  $F_0$ -Verlauf ist und selbst nicht durch linguistische Parameter geprägt ist, weshalb es stärker kategorisierenden und phonologisch interpretierenden Verfahren wie ToBI vorgezogen worden ist.

### 1.3 Umsetzung

An vier Orten wurden Aufnahmen mit je ca. 20 Maturanden (Abiturienten) durchgeführt. Die vier Städte sind je unterschiedlichen Mundartregionen des Alemannischen zuzuordnen: Brig im Wallis (WS) gehört zum Höchstalemannischen im alpinen Südwesten. Chur in Graubünden (GR) ist wie die anderen Orte dem Hochalemannischen zuzuordnen und liegt im östlichen Alpenraum. Bern, Hauptstadt der Schweiz und des Kantons Bern (BE), ist das Zentrum der westlichen Mundarten im Schweizer Mittelland, während Winterthur, die zweitgrößte Stadt im Kanton Zürich (ZH) die zentral-östlichen Teile des Mittellandes vertritt. Die ursprünglich geplanten Aufnahmen im niederalemannischen Basel, in der Nordostschweiz mit St. Gallen und in der höchstalemannischen Innerschweiz mit Schwyz mussten aus Kapazitätsgründen fallen gelassen werden. Für die Aufnahmen hatten die Schulleitungen jeweils ein ruhiges Zimmer zur Verfügung gestellt. Die Interviews wurden mit Richtmikrofonen digital aufgenommen. Von den Aufnahmen wurden jeweils fünf von Männern und fünf von Frauen für die weitere Analyse ausgewählt, wobei die dialektale Sicherheit der Sprecher\_innen, deren Klarheit der Stimme sowie die akustische Qualität der Aufnahme als Auswahlkriterien angesetzt wurden.

Für die statistische Analyse wurden die Aufnahmen auf Lautebene segmentiert. Dazu wurden von diesen Aufnahmen je gut drei Minuten Sprechzeit mit einem für die Schweizer Mundarten angepassten SAMPA-Transkriptionssystem transkribiert und mit einem an der Universität Lausanne im Vorgängerprojekt entwickelten Aligner auf Lautebene vorsegmentiert. Anschließend wurden die automatisch gesetzten Labels mittels Praat (Boersmaa & Weenink 2014) manuell anhand eines Annotationsschemas korrigiert. Notiert wurden für jeden Laut die phonetische Realisierung und die unterliegende phonologische Vollform, wie das in Abbildung 2 dargestellt ist.



Sichtbar wird in diesem Ausschnitt die Markierung der Reduktion von [a] (\*<a), die Komplettassimilation von [d] und [ts] (\_ts<Vd\_ts) bzw. (ts<dts) sowie daselbst Trennung von Okklusion und Burst bei Plosiven und Affrikaten. Ebenso die Komplettassimilation von [d] und [k] (C<dk) über die Wortgrenze hinweg. Insgesamt sind so etwas über 123000 Laute manuell segmentiert worden.

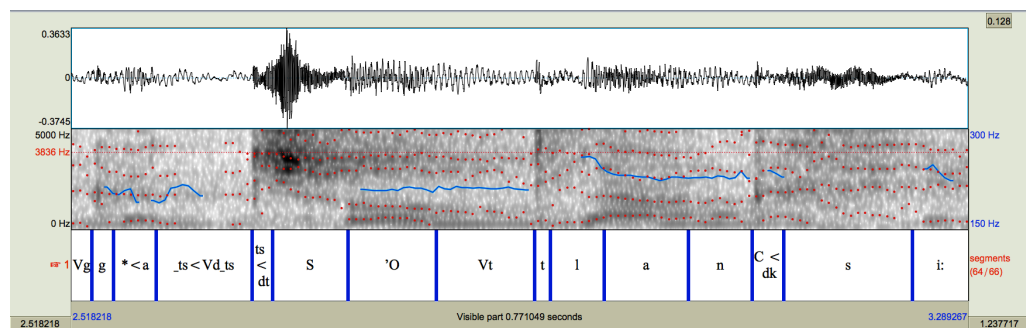


Abbildung 2: Beispiel für die Segmentierung auf Lautebene: ... *gad z schottland gsii* – ... gerade in Schottland gewesen

Um eine einigermaßen einheitliche Segmentierung zu gewährleisten, wurden die ersten Aufnahmen integral von mindestens zwei Mitarbeiter\_innen gelabelt. Sämtliche Labels, die sich um mehr als 10 ms unterschieden, wurden dann in der Gruppe gemeinsam analysiert. Unsicherheiten mit der Segmentierung konnten so grundsätzlich auf ein Minimum reduziert werden. Einige Fälle (typischerweise das Ende einer Phrase vor einer Pause; Grenzen zwischen Vokalen oder zwischen Frikativen) sind jedoch intrinsisch problematisch und sind bis zum Ende des Projekt problembehaftet geblieben. Im Verlauf der ganzen Segmentationszeit sind solche Vergleiche für Teile der Aufnahmen regelmäßig durchgeführt worden und die Mitarbeiter\_innen haben bei Unsicherheiten immer andere Meinungen eingeholt, so dass die Annotationen

zwischen den Labellern und auch beim selben Labeller über die Zeit stabil blieben.

Die auf diese Weise erstellten Labels mit Angabe der Lautdauer, der Lautqualität und der phonologischen Grundlage wurden für die linguistische Annotation und die anschließende statistische Analyse in eine Tabelle übertragen. Die einzelnen Segmente sind in der Annotation mit den oben genannten linguistischen und paralinguistischen Informationen ergänzt worden. Insbesondere wurden Informationen zur Silbengrenze, zur lautlichen Umgebung, zu Akzentsetzung und Fokus, zur Phrasenstrukturierung und zum lexikalischen Status der zugrundeliegenden Wörter angefügt. Da die Aufnahmen besonders für die Bestimmung der Phrasengrenzen und von Fokus nochmals angehört werden mussten, ist gleichzeitig eine Korrektur der Transkription erfolgt.

Mehrere der annotierten Parameter sind stark theorieabhängig, insbesondere die Definition von Silben und Phrasen. Deren Definition prägt das Resultat der Studie, weshalb an dieser Stelle nochmals darauf eingegangen wird. Da das Projekt eine phonetische Ausrichtung hat, sind die Definitionen eher an den phonetischen Kriterien orientiert. Für die Bestimmung der Silbengrenzen war z. B. nur die Sonoritätshierarchie bestimmend, die für die Schweizer Mundarten auch in der Literatur höher gewertet wurde als Wortgrenzen. Dieser Entscheid ist durch die häufigen Sandhi-Erscheinungen im Schweizerdeutschen (Moulton 1986) geprägt und zudem durch die typologische Betonung silbensprachlicher Züge der Deutschschweizer Mundarten (Nübling & Schrambke 2004; Siebenhaar 2014) bestärkt worden. Die Modellierung des Timings hat dann auch gezeigt, dass die Berücksichtigung der Wortgrenze die Modelle nicht verbessert und deshalb vernachlässigt werden konnte.

Die Markierung der Phrasen (IP) erfolgte auf der Basis ihrer konversationsstrukturierenden Funktion als weiterführende oder abschließende Phrasen, wie das von Gilles (2005) vorgeschlagen wurde. Zudem wurden Fragephrasen oder unterbrochene Phrasen markiert. Potenzielle Phrasengrenzen, welche durch die Abfolge eines lexikalischen und eines folgenden grammatischen Wortes definiert sind (Grosjean & Collins 1979, Siebenhaar, Zellner Keller & Keller 2001) und sich als iP meist durch eine phrasenfinale Dehnung ohne Pause auszeichnen, wurden ebenfalls markiert.

Um die Parameter für die Intonationsverläufe auszuarbeiten, wurden die Grundfrequenzverläufe zuerst mittels Praat (Boersmaa & Weenink 2014) ausgemessen und Fehlmessungen manuell korrigiert. Die korrigierten Frequenzverläufe sind dann mittels Mixdorffs (2009) Programm mathematisch beschrieben worden und die Kommandos sind den Phrasen bzw. akzentuierten Silben zugeordnet worden, wobei Akzent in diesem Fall nur als Intonationsakzent zu verstehen ist. Die Unterscheidung von lexikalischem Wortakzent (*stress*) und Intonationsakzent (*accent*) ist hier bedeutsam, weil sich die Intonationsakzente nicht nur auf den Wortakzent beziehen lassen. Im Verlauf der Arbeit zeigt sich nämlich immer deutlicher, dass – besonders im Walliser Dialekt – Intonationsakzente auch auf Nebensilben und sogar auf unbetonten Schwa-Silben liegen können.

Die Analysen und Modellierungen sind für Timing und die Intonation unabhängig durchgeführt worden. Durch die unterschiedlichen Ankereinheiten – Segment bzw. Silbe – ist die ursprünglich geplante gemeinsame Analyse nicht mehr durchgeführt worden. Bezüge sind nur interpretativ hergestellt worden. Als Analysen sind ANOVAs und Regressionsanalysen bzw. Generalisierte Lineare Modelle eingesetzt worden.

## 2 Analysen / Ergebnisse

### 2.1 Phrasierung

Ein wesentlicher regionaler Unterschied in der Prosodie zeigt sich in der Strukturierung von Aussagen, der Phrasierung: Zürcher realisieren mit einer durchschnittlich Phrasenlänge von 6.4 Silben die längsten Phrasen. Dann folgen die Bündner mit 5.4 Silben / Phrase und Walliser und Berner mit je 5.0 Silben / Phrase.<sup>2</sup> Damit manifestiert sich ein signifikanter West / Ost-Unterschied, der sich vor allem mit der Häufung ganz kurzer Phrasen mit einer oder zwei Silben in den westlichen Mundarten (BE und WS) erklären lässt. Dieses Verhalten kann nicht abschließend geklärt werden, der Eindruck erhärtet sich aber, dass diese kurzen Phrasen v. a. einleitend als Operatoren (im Sinne der Operator-Skopus-Strukturen, Fiehler et al. 2004) oder manchmal als *tag questions* oder kurze Wiederholungen vorkommen. Somit kann vermutet werden, dass sich der prosodische Unterschied als Folge eines unterschiedlichen Kommunikationsverhaltens erklären lässt. Ein solcher kultureller Unterschied müsste aber in weiterführenden Untersuchungen mit anderen (ethnomethodologisch-gesprächslinguistischen) Methoden erhärtet werden.

Auch die Anzahl der Pausen und deren Dauer zeigen eine unterschiedliche Verteilung zwischen den Dialekten. In allen vier Dialekten sind abschließende Phrasen normalerweise von einer Pause gefolgt. Nach weiterführenden Phrasen

<sup>2</sup> Für die Intonationsanalyse musste eine Anzahl Phrasen ausgeschlossen werden, da der Intonationsverlauf besonders bei sehr kurzen Phrasen zu wenig deutlich oder durch Knarrstimme schlecht zu analysieren war. Entsprechend unterscheiden sich die Daten in Leemann (2012: 174–175) von den hier und auch auch in Leemann / Siebenhaar (2008a) dargestellten, wobei die Grundtendenz dieselbe ist.

verhalten sich die Walliser Sprecher aber anders als die Sprecher der anderen Mundarten: Sie zeigen signifikant weniger Pausen und da, wo sie doch Pausen setzen, sind diese signifikant kürzer als die der anderen Sprecher. Die Walliser Sprecher zeigen damit eine sehr hohe Sprechgeschwindigkeit und unterscheiden sich von den Zürchern, die eine ähnlich hohe Artikulationsgeschwindigkeit, aber längere und auch mehr Pausen aufweisen.

## 2.2 Timing

Die zeitliche Strukturierung der vier Dialekte unterscheidet sich in vielerlei Hinsicht. Hier werden einzelne Aspekte herausgegriffen, welche die rhythmische Gestaltung prägen. Als erstes zeigt sich, dass die vier Mundarten eine je andere Verteilung der Nukleitypen aufweisen. Tabelle 1 listet die Nuklei betonter Silben auf und zeigt, dass sich die Verhältnisse zwischen den Mundarten um einige Prozente unterscheiden. Während also in der Walliser Mundart 66 % aller Vokale phonologische Kurzvokale sind, entspricht deren Anteil in der Bündner Mundart nur 57.5 %.

Mundart / %	Diphthonge	Langvokale	Kurzvokale	%; N
BE	16.5 %	19.7 %	63.8 %	100 % =3241
ZH	15.0 %	23.0 %	62.0 %	100 % =3334
WS	16.6 %	17.3 %	66.0 %	100 % =3235
GR	18.8 %	23.7 %	57.5 %	100 % =3258
N	2184	2739	8145	13068

Tabelle 1: Anteil wortakzentuierter Diphthonge, Langvokale und Kurzvokale nach Aufnahme-region

Abbildung 3 macht zudem deutlich, dass sich nicht nur die Anzahl der jeweiligen Nukleitypen unterscheiden, sondern auch deren durchschnittliche

Dauer. Während das durchschnittliche Dauerverhältnis zwischen Langvokalen und Kurzvokalen für alle Mundarten ähnlich ist, so ist deren Verhältnis zu den Diphthongen jeweils anders: Während die westlichen Dialekte (BE, WS) die Diphthonge im Verhältnis zu den Langvokalen kürzer realisieren, sind sie in den östlichen Mundarten (ZH, GR) länger, wobei der Unterschied zwischen Langvokalen und Diphthongen außer bei den Bündnern immer signifikant ist. Aus der Abbildung wird auch deutlich, dass Zürcher und Walliser im Schnitt aller Vokale deutlich kürzere Lautdauern aufweisen als Berner und Bündner. Diese Gruppen unterscheiden sich also signifikant in der Artikulationsgeschwindigkeit, während innerhalb der beiden Gruppen keine signifikanten Unterschiede bestehen. Die oben aufgestellte West/ Ost-Differenz trifft also nur auf die Verhältnisse zwischen den Lauttypen zu, im Bezug auf die Gesamtdauer der betonten Vokale gilt sie nicht.

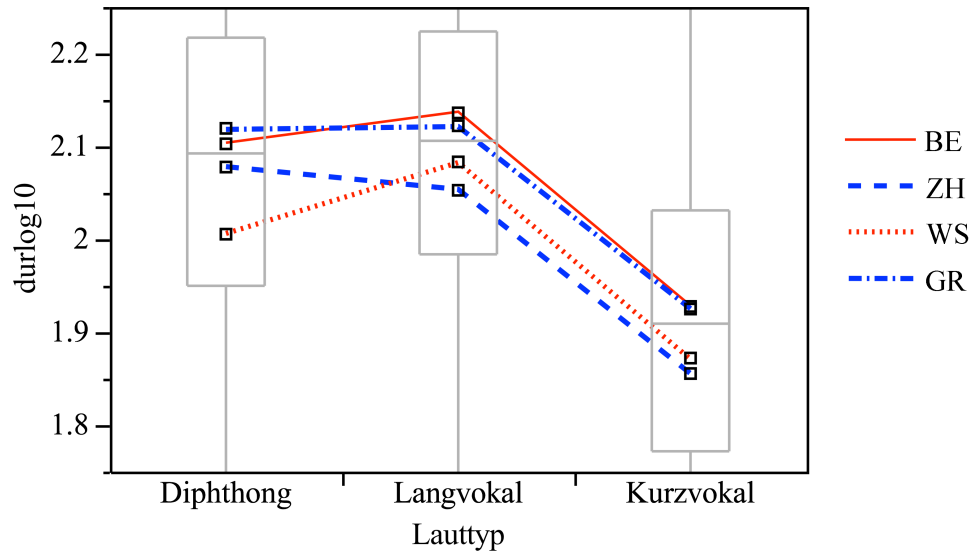


Abbildung 3: Dauer der wortakzentuierten Vokale nach Lauttyp (Diphthong, Langvokal, Kurzvokal) und nach Aufnahme-region

Abbildung 4 zeigt den Unterschied der Silbenkerne in Silben, die nicht durch die Phrasengrenzen gedehnt sind (s. u.). Wieder wird die Dauerstaffelung von  $WS \leq ZH < GR < BE$  deutlich, wobei die Unterschiede zwischen den Regionen außer zwischen ZH und WS signifikant sind. Die Unterschiede zwischen den *stress*-Typen innerhalb der Mundarten sind alle signifikant. Es zeigen sich aber auch unterschiedliche Dauerverhältnisse zwischen den *stress*-Typen: Während Berner und Walliser Schwa-Vokale, Vollvokale ohne Wortakzent und Vollvokale mit Wortakzent jeweils ähnlich voneinander unterscheiden, zeigen Zürcher und Bündner einen deutlichen Unterschied zwischen den Vollvokalen mit und ohne Wortakzent, während sie Schwa-Silben weniger deutlich von den unbetonten Vollvokalen unterscheiden.

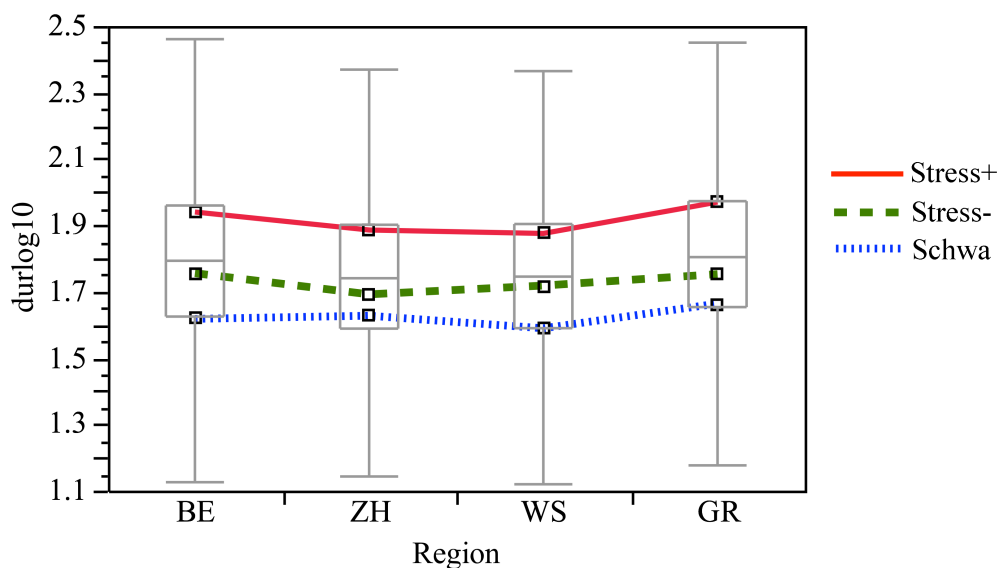


Abbildung 4: Dauer der Nuklei in phrasenmittigen Silben nach Aufnahme-region und nach *stress* (Stress+: Wortakzent; Stress-: Vollvokale ohne Wortakzent; Schwa-Silben)

Abbildung 5 zeigt Unterschiede im Timing, welche die Unterschiede der Phrasierung auf der Timingebene abbilden. Die Abbildung gibt die Dauer der Nuklei nach Position in der Silbe in der Phrase wieder. Unterschieden werden die erste Silbe (f), die mittleren (m), die vorletzte (p) und die letzte (u) Silbe. Phrasenenden werden in allen Mundarten mit einer Dehnung der letzten beiden Silben markiert. Auch der Phrasenbeginn wird mit einer Dehnung der ersten Silbe hervorgehoben. Alle Dauerunterschiede sind dabei signifikant, außer die Dehnung der ersten Silbe gegenüber den mittleren bei den Wallisern. Die Walliser unterscheiden sich aber nicht nur in der ersten Silbe der Phrase von den übrigen Sprechern, sondern besonders auch in der Phrasenenddehnung, insofern diese Dehnung weit weniger ausgeprägt ist. Wenn man noch weiter in die Details geht, so äußert sich dieser Unterschied darin, dass die letzten Silben in weiterführenden Phrasen deutlich stärker gedehnt werden als in



abschließenden Phrasen. Dieser Unterschied zwischen weiterführend und abschließend ist bei den Wallisern aber nicht signifikant und bei den Bernern weniger deutlich als in den östlichen Mundarten.

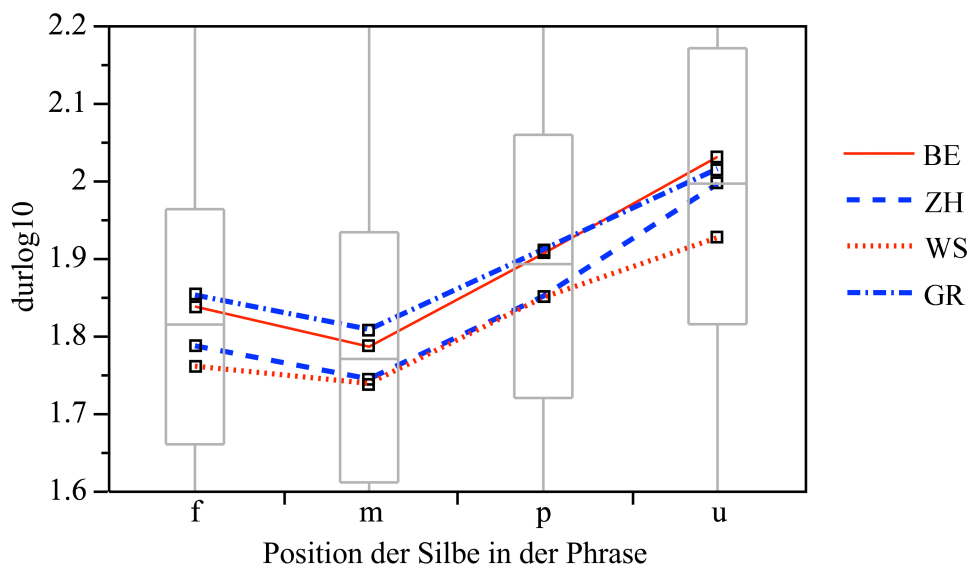


Abbildung 5: Dauer der Nuklei aller Silben nach Position der Silbe in der Phrase (erste, mittlere, vorletzte, letzte) und nach Aufnahme-region

Die Vielzahl von ANOVAs zeigt, dass Einflüsse auf die Dauer der einzelnen Laute überaus vielfältig und in sehr vielen Bereichen auch zwischen den Mundarten signifikant unterschiedlich gestaltet sind. Diese Unterschiede zwischen den Mundarten werden aber einerseits durch individuelle Variation überlagert. Andererseits sind auch die Interaktionen zwischen den einzelnen Einflussfaktoren sehr komplex. Dadurch werden Differenzen einzelner Variablen überdeckt, wodurch sie in einer ANOVA nicht erkannt werden. Es kann aber auch vorkommen, dass eine Variable überbewertet wird, weil deren Einfluss durch andere Variablen, welche selbst nicht in die Einzelanalyse

einfließen, gestützt wird. So ist beispielsweise die ursprünglich mitberücksichtigte Kategorie Wortklasse, die grammatische von lexikalischen Wörtern unterscheidet, im Verlauf der Arbeit fallen gelassen worden, da sie in der Modellierung der beiden Alpinmundarten keinen signifikanten Einfluss zeigt und in Bern und Zürich nur einen sehr geringen. Der in allen Dialekten hochsignifikante prosodische Unterschied, der durch Wortklasse bedingt erscheint, wird also mehrheitlich durch andere Faktoren abgedeckt. Um diese Komplexität der Interaktionen abzubilden, wird für jede Mundart ein General Linear Model (GLM) erstellt ([Leemann & Siebenhaar 2010](#)). Ein solches Modell berücksichtigt eine (theoretisch unendliche) Anzahl von Variablen, die mit einem je unterschiedlichen Faktor in das Modell einfließen. Mit dem Vergleich der Modelle kann die Gewichtung der einzelnen Variablen nachgezeichnet werden. Für die Modellierung des Timings, der Dauer jedes einzelnen Segments, werden schließlich die folgenden Variablen berücksichtigt: intrinsische Lautdauer, Position des Lauts in der Silbe, Typ des vorangehenden und des folgenden Segments, Totalassimilation von Plosivfolgen, Fokus und *stress*, Position der Silbe in der Phrase. Die folgende Tabelle 2 zeigt, wie gut (bzw. wie schlecht) die Variation der einzelnen Mundarten erklärt werden kann.

Mundart	BE	ZH	WS	GR
Erklärte Varianz	47.9 %	46.5 %	44.1 %	53.1 %

Tabelle 2: Anteil der Varianz, der durch die GLMs erklärt wird ([Leemann & Siebenhaar 2010](#))

Tabelle 2 offenbart, dass die Modelle nur knapp die Hälfte der gesamten Varianz erklären können, wobei das Bündner Modell sehr deutlich besser ist

als die der anderen Mundarten. Das Timing der Bündner Sprecher kann daher als das regelmäßigste gelten. Im Gegenzug ist das Timing der Walliser viel variabler. Im Vergleich zu diesen Modellen der Dialekte erklären Modelle zur Lesesprache von Einzelpersonen fast zwei Drittel der gesamten Variation (Siebenhaar, Zellner Keller & Keller 2001). Auch die Modelle zur dialektalen Spontansprache von Einzelpersonen erreichen deutlich höhere Werte (Häsler, Hove & Siebenhaar 2005). Modelle für die Individuen der vorliegenden Datenbasis zeigen mit wenigen Ausnahmen auch eine größere Erklärungsadäquatheit. Diese Modelle für Individuen werden jedoch nicht systematisch untersucht, weil sie wegen zu kleiner Datenbasis oft überspezifiziert sind. Damit zeigt sich auch in der Modellierung, dass die individuelle Variation die dialektale Variation überlagert. Trotzdem lassen sich dialektale Unterschiede deutlich erkennen, wie aus Tabelle 3 ersichtlich wird, in welcher das relative Gewicht der einzelnen Faktoren im Modell dargestellt wird. Dabei sind nur diejenigen Variablen berücksichtigt, welche in allen Mundarten einen signifikanten Einfluss ausüben.

	BE	ZH	WS	GR
Intrinsische Dauer (12 df)	57.8 %	46.4 %	57.9 %	56.2 %
Position der Silbe in der Phrase (4 df)	16.3 %	21.7 %	15.7 %	13.5 %
Fokus und <i>stress</i> (4 df)	9.7 %	13.1 %	12.6 %	13.4 %
Typ des folgenden Segments (12 df)	8.7 %	9.1 %	8.7 %	8.1 %
Typ des vorangehenden Segments (12 df)	4.4 %	6.0 %	2.5 %	5.2 %
Position des Segments in der Silbe (3 df)	2.5 %	2.6 %	1.2 %	1.7 %
Totalassimilation doppelter Plosive (4 df)	0.6 %	1.1 %	1.4 %	1.8 %

Tabelle 3: Gewichtung der einzelnen Variablen im GLM mit Angabe der Freiheitsgrade (df) (Leemann & Siebenhaar 2010)

Der Vergleich offenbart, dass die Reihenfolge der Gewichtung der Faktoren bei allen vier Mundarten dieselbe ist. Was die vier Mundarten unterscheidet, ist

das relative Gewicht der Faktoren. So ist auffällig, dass die intrinsische Dauer bei den Zürchern eine geringere Rolle spielt, die Variation daher stärker auf die Position in der Phrase und auf Fokus und *stress* zurückzuführen ist. Die Berner zeigen im Gegenzug weniger Einfluss von Fokus und *stress*. Für die Modellierung wird die Variable Wortklasse nicht berücksichtigt, weil sie in den Alpinmundarten keinen signifikanten Einfluss zeigt. Die Position der Silbe in der Phrase ist für die mittelländischen Mundarten (BE, ZH) etwas wichtiger als für die Alpinmundarten (WS, GR), d. h. die Phrasenenddehnung ist in den mittelländischen Mundarten deutlicher ausgeprägt als in den Alpinmundarten, wobei die östlichen Mundarten (ZH, GR) je etwas stärkere Dehnungen aufweisen als die westlichen (BE, WS). Als deutlichster West / Ost-Unterschied erweist sich die intrinsische Dauer. Das ist u. a. auf die Dauerunterschiede der Diphthonge zurückzuführen, die im Westen jeweils kürzer sind als die Langvokale, während sie in den östlichen Mundarten länger sind als die Langvokale. Da die intrinsische Dauer der gewichtigste Faktor im Timing darstellt, ist dieser Unterschied als sehr bedeutend zu werten. Auch der *stress*-Unterschied zeigt eine West / Ost-Differenzierung, da die östlichen Mundarten deutlicher zwischen wortakzentuierten und nicht-wortakzentuierten Vollvokalen unterscheiden als die westlichen Mundarten, die dagegen die Schwa-Silben deutlicher von den Vollvokalen ohne Wortakzent unterscheiden. Die Mundarten unterscheiden sich auch in der durchschnittlichen Dauer der Laute und damit der Artikulationsgeschwindigkeit. Dabei ist keine West / Ost- oder Nord / Süd-Unterscheidung anzusetzen: Die kürzesten Laute zeigen die Walliser, gefolgt von den Zürchern, den Bündnern und Bernern, die als langsamste Sprecher gelten, was mit dieser Untersuchung auch empirisch bestätigt werden konnte.

### 2.3 Intonation

Die Intonation wurde detailliert von Adrian Leemann (2012) beschrieben; die folgenden Ausführungen stützen sich im Wesentlichen auf diesen Text. Generell ist festzuhalten, dass die Variation in der Intonation als viel komplexer erscheint als im Timing, da auch mit der Fujisaki-Parametrisierung jeweils mehrere Variablen modelliert werden müssen. Die Intonation ist dadurch noch mehr als das Timing durch viele Faktoren spontan gesprochener Sprache mitbestimmt, welche in der Untersuchung nicht berücksichtigt sind, wie die inhaltliche und konversationelle Strukturierung, *disfluencies*, Wiederholungen u. v. m. Der Versuch, für die einzelnen Dialekte Modelle zu entwickeln, macht deutlich, dass die einer Äußerung zugrundeliegenden linguistischen Parameter wie Silbentyp, Wortart oder *stress* für die Intonation weniger wichtig sind als die paralinguistischen Parameter wie Fokus, prosodische Abschnittmarkierung, Phrasentyp, welche vom Sprecher bewusst eingesetzt werden um eine Äußerung zu strukturieren. Damit zeigt sich ein deutlicher Unterschied zu Modellierung des Timings, wo Abschnittmarkierung nicht und Phrasentyp nur als beschränkt relevant erscheinen und die linguistischen Parameter deutlich mehr Gewicht einnehmen. Die Intonationsmodelle erklären dadurch relativ geringe Anteile der gesamten Varianz. Als besonders schwierig erweist sich der statistische Vergleich von Modellen, da jede einzelne Mundart unterschiedliche Einflüsse anders gewichtet, wodurch gemeinsame Modelle, die nur signifikante Variablen aufgreifen, ein reduziertes Erklärungspotenzial haben. Dieser Vergleich ist deshalb weniger statistisch, sondern eher ein interpretativer Vergleich von dialektspezifischen Besonderheiten, die statistisch abgestützt sind.

Für die Analyse sind die folgenden Faktoren als Erklärung intonatorischer Variation berücksichtigt worden: phonologische Länge des Silbenkerns, Silbenstruktur, *stress* und Wortklasse als linguistische Variablen. Als paralinguistische: Fokus, Phrasentypus und prosodische Abschnittmarkierung. Als nicht-linguistische: Sprechgeschwindigkeit, Emotionalität und Geschlecht. Die globalen Variablen der Intonation, also Phrasenkommandoposition und -höhe, lassen sich generell etwas besser erklären als die lokalen Akzentkommandos – Position, Höhe und Dauer. Für die Phrasenkommandos erweisen sich besonders die Bezüge zu den vorangehenden Phrasen als bedeutsam, was als Einfluss der Informationsstrukturierung gewertet werden kann. Die Mundarten unterscheiden sich relativ wenig. Für die Erklärung der Akzentkommandos sind Phrasentyp und Fokus die beiden wesentlichen Faktoren. Beide sind ebenfalls für die Informations- bzw. Interaktionsstrukturierung bedeutsam, was einerseits erklärt, dass ihre Gewichtung zwischen den Dialekten keine Unterschiede zeigen. Andererseits bestätigt das auch die Annahme, dass zwischen den Mundarten nicht phonologische, sondern phonetische Intonationsunterschiede bestehen. Nichtsdestoweniger zeigen sich auch Unterschiede zwischen den Dialekten, die im Folgenden erläutert werden.

Die folgenden Abbildungen 6–9 stehen exemplarisch für die Intonation der Mundarträume. Die Abbildungen zeigen jeweils mit Kreuzen in der dicken Linie dargestellt die gemessene Grundfrequenz. Eine ausgezogene Linie folgt diesem Verlauf möglichst genau. Dies ist die mit dem Fujisaki-Modell generierte F<sub>0</sub>-Kontour, wie sie schon in Abbildung 1 schematisch dargestellt wurde. Diese Kontour setzt sich zusammen aus den relativ langsamen Phrasenkommandos, die als untere Linie bogenförmig einzelne Phrasen darstellt, und den darauf aufbauenden schnellen Akzentkommandos. Der

Punkt, an dem die Modelllinie jeweils wieder nach oben ausbricht, ist die Stelle des ersten Akzentkommandos, und da, wo ein Fall wieder deutlich wird, ist das zweite Akzentkommando. Die Phrasenkommandos werden in der linguistischen Interpretation mit dem Phrasenbeginn, die Akzentkommandos mit Silben verbunden.

Der Berner Dialekt ist gekennzeichnet durch späte Akzentkommandos, d. h. der  $F_0$ -Anstieg der lokaler Akzente beginnt meist erst nach dem Silbenanfang, mit dem das Akzentkommando verbunden ist, das  $F_0$ -Maximum des Akzents wird oft erst in der nächsten Silbe erreicht. Diese Akzentkommandos sind zudem meist mit lexikalischem *stress* verbunden. Berner haben auch relativ lange Akzentkommandos, was mit der relativ geringen Artikulationsgeschwindigkeit zusammenhängt, die schon in der Timinganalyse als lange Nuklei und als deutliche phrasenfinale Dehnung offenbar wurde. Abbildung 6 zeigt die typischen späten  $F_0$ -Anstiege, die besonders in den drei eingekreisten Stellen sehr deutlich werden, wo das Akzentkommando nach dem Beginn der Silbe einsetzt und das  $F_0$ -Maximum erst in der nächsten Silbe erreicht wird.

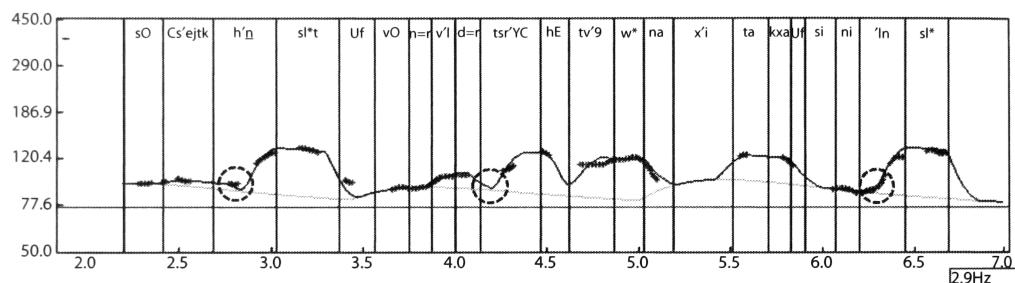


Abbildung 6: Typische Berner Intonationskontur: Späte Anstiege in lokalen Akzentkommandos – *so gseit ghänslet [uf], won er wider zrügget het wöwwe nach Itaka uf sini Insle.* – So gesagt gehänselt [auf], als er wieder zurück hat wollen nach Ithaka auf seine Insel (Leemann 2012: 276)

Der Zürcher Dialekt zeigt eher hohe Phrasenkommandos, was mit der Tendenz zu längeren Phrasen zusammenhängt, zudem ist die Höhe des Phrasenkommandos von der Länge der vorangehenden Pause abhängig. Spezifisch für die Zürcher Sprecher\_innen ist, dass nur bei ihnen Akzentkommandos mit der Wortklasse korrelieren, d. h. Akzentkommandos kommen bei grammatischen Wörtern kaum vor. Abbildung 7 zeigt ein langes Phrasenkommando, das über mehrere syntaktisch unvollständige iPs hinausreicht; zudem wird deutlich, dass die lexikalischen Wörter alle mit Akzentkommandos versehen sind.



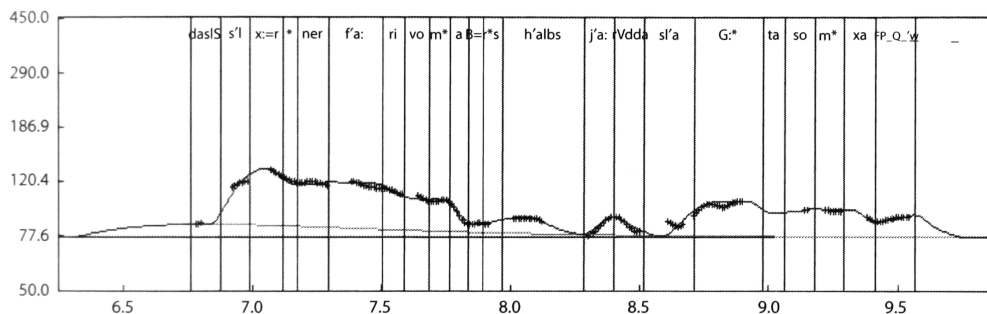


Abbildung 7: Typische Zürcher Intonationskontur: lange IPs – *Das isch sicher en Erfaarig, wo me aber es halbs Jahr, das langet, also me cha au...* – Das ist sicher eine Erfahrung, wo man aber ein halbes Jahr, das reicht, also man kann auch... (Leemann 2012: 283)

Der Walliser Dialekt zeigt generell die schlechteste Erklärungsadäquatheit mit den untersuchten Faktoren. Gegenüber den anderen Dialekten sind die Faktoren deutlich unterschiedlich gewichtet. So sind Akzente vor fokussierten Silben weniger reduziert als in anderen Dialekten, die fokussierten Silben sind dagegen weniger deutlich hervorgehoben. Die letzten Akzentkommandos in abschließenden Phrasen sind im Vergleich mit denen der anderen Mundarten sehr klein. Zudem sind die Akzentkommandos nicht mit der Wortklasse oder *stress* verbunden, d. h. Akzentkommandos können mit jeder, auch einer Schwa-Silbe eines grammatischen Wortes verbunden werden und sind von der Höhe her immer wieder unterschiedlich. Insgesamt korrelieren sowohl Akzent- wie auch Phrasenkommandos gleichzeitig mit vielen der untersuchten Faktoren. Die Walliser Intonation ist damit sehr komplex, was die Vorhersage schwierig macht. Abbildung 8 zeigt deutlich, dass die kurzen lokalen Akzentkommandos immer unterschiedliche Höhe aufweisen und mit verschiedenen Silbentypen verbunden sind, was einer sehr lebhaften  $F_0$ -Kontur entspricht.

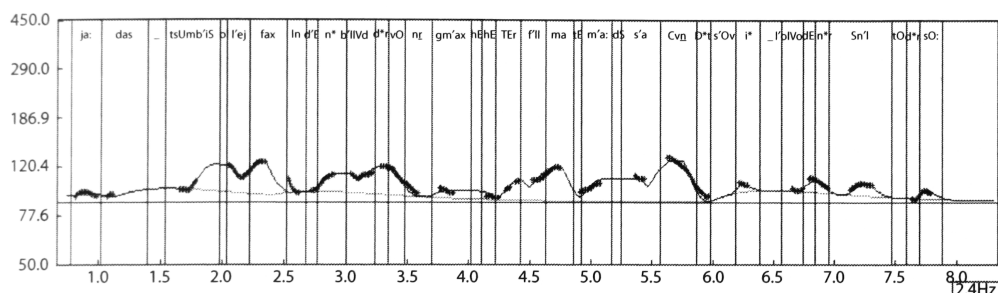


Abbildung 8: Typische Walliser Intonationskontur: Unterschiedliche lokale Akzentkonturen – *Jaa das, zum Bispil einfach in dene Bilder von er gmacht het, het er vil Matemaatisches aaggwändet, so wie goldener Schnitt oder soo* – Ja das, zum Beispiel einfach in diesen Bildern, die er gemacht hat, hat er viel Mathematisches angewendet, so wie goldener Schnitt oder so. (Leemann 2012: 281)

Auffällig sind für den Bündner Dialekt relativ flache hohe  $F_0$ -Konturen. Diese zeigen ein generell hohes Phrasenkommando, das nur von der Dauer der vorangehenden Pause abhängig ist; die flache Kontur wird im Modell mit ansteigenden Akzentkommandos wiedergegeben, was sich nur im Bündner Dialekt zeigt. Gegenüber den anderen Dialekten sind deutlich mehr fallende Akzente festzustellen. Wie bei den Wallisern sind die Akzentkommandos nicht mit Wortklasse oder *stress* verbunden. Trotzdem ist die Modellierung der Bündner Intonation die beste aller vier Mundarten, was v. a. darauf zurückzuführen ist, dass bei der Voraussage der einzelnen Variablen die meisten Faktoren wenig interagieren. Abbildung 9 zeigt eine typische flache hohe  $F_0$ -Kontur mit fallenden Akzenten auf *Primaarleererin* und *do*.

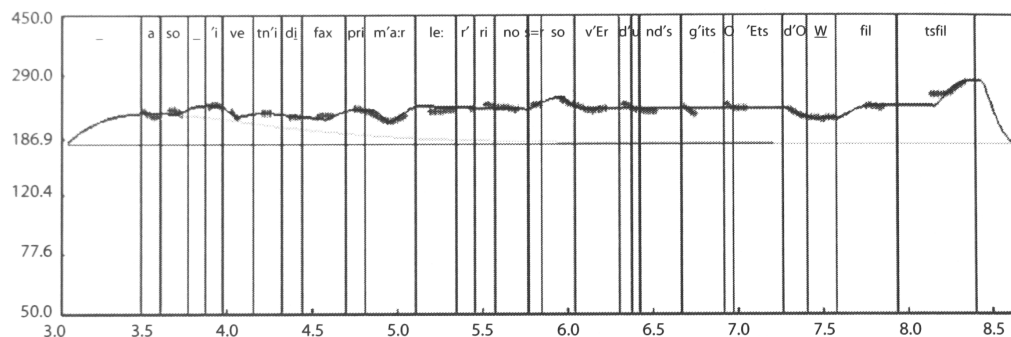


Abbildung 9: Typische Bündner Intonationskontur: hohe, flache  $F_0$ -Kontur – *Aso, i wet nid äifach Primaarleererin oder so werde und es gits jo jetzt do au fil z fil.* – Also ich will nicht einfach Primarlehrerin oder so werden und es gibt [es] ja jetzt da auch viel zu viele. (Leemann 2012: 278)

Insgesamt zeigen sich Nord / Süd-Gegensätze als Unterschied des Einflusses von Wortklasse und *stress*, welche in den mittelländischen Mundarten (BE, ZH) einen Einfluss zeigen, der in den Alpinmundarten (GR, WS) nicht vorhanden ist. Tendenziell zeigt sich in den mittelländischen Mundarten, insbesondere in Bern ein später Einsatz des Akzentkommandos. West / Ost-Gegensätze finden sich eher in der Bedeutung der Phrasenkommandos, die im Westen bei weiterführenden Phrasen deutlich höher sind als bei abschließenden. Dieser Unterschied wird in den östlichen Mundarten kaum gemacht. Dort sind die Phrasenkommandos aber deutlich länger als in den westlichen Mundarten.

#### 2.4 Timing und Intonation

Der Vergleich von Timing und Intonation offenbart eine relative Unabhängigkeit der beiden prosodischen Aspekte. Gewisse Parallelen zeigen sich in der Markierung der Phrasengrenzen, die in den östlichen Mundarten im

Timing und in der Intonation deutlicher ist als in den westlichen Mundarten. Bei der Unterscheidung von weiterführenden und abschließenden Phrasen scheinen sich dagegen Timing und Intonation gegenseitig zu kompensieren. Während die östlichen Mundarten den Unterschied mit dem Timing markieren, indem die Schlusssilben in weiterführenden Phrasen deutlicher gedehnt werden als in abschließenden, markieren ihn die westlichen Mundarten stärker mittels der Intonation, indem die letzten Akzentkommandos in abschließenden Phrasen viel stärker reduziert sind als in weiterführenden. Die Faktoren Wortakzent (*stress*) und Wortklasse sind im Timing und in der Intonation in den mittelländischen Mundarten bedeutender als in den Alpinmundarten. Als Erklärung für diese Unterschiede kann die Nähe zu den romanischen Sprachen angeführt werden. Ebenso ist es möglich als Ursache für die Unterschiede eine archaischere Prosodie anzuführen, die sich als stärkere Silbensprachlichkeit der südlicheren Mundarten äußert (vgl. Szczepaniak 2007, Siebenhaar 2014), wobei letzteres sich mit den tendenziell archaischeren Zügen auf segmentaler Ebene deckt (vgl. Hotzenköcherle 1984).

## 2.5 Perzeption

Der Schwerpunkt im Projekt liegt auf der Erfassung akustischer Unterschiede im Bereich der Prosodie. Im Verlauf der Arbeit stellte sich allerdings auch die Frage, inwiefern die gemessenen Unterschiede überhaupt wahrgenommen werden (Leemann & Siebenhaar 2008b). Obwohl sich Deutschschweizer stark über ihren Dialekt geographisch verorten (Werlen 2005) und die Prosodie häufig als für gewisse Dialekte charakteristisch bezeichnet wird (u. a. schon Stalder 1819: 8, Ris 1992), ist noch offen, ob für eine sprachgeographische Zuordnung innerhalb der Schweizer Dialekte die Prosodie allein genutzt

werden kann. Um dies zu überprüfen, wurden aus jeder der vier regionalen Sprechergruppen zwei männliche Sprecher ausgewählt, die im Bezug auf die beiden untersuchten prosodischen Aspekte als typisch gelten können. Aus deren Aufnahmen wurden von jedem drei vergleichbare, je 5–7 Sekunden lange Ausschnitte ausgewählt. Um perzeptionsrelevante Faktoren außer der Prosodie auszuschließen, wurden zwei der Aufnahmen delexikalisiert. Dazu wurden sie zuerst auf eine gleiche mittlere Grundfrequenz von 140 Hz gesetzt. Anschließend wurden die Frequenzen über 250 Hz mit einem Stop-Hann-Band-Filter gefiltert und bei 500 Hz geglättet. Zum Schluss wurden die Amplituden auf eine einheitliche Lautstärke gebracht. Nach diesem Procedere haben alle Texte die gleiche mittlere Grundfrequenz und sie sind gleich laut. Die Texte sind nicht mehr verständlich und klingen so dumpf, wie wenn man ein Gespräch durch eine Wand hören würde.

Für das Perzeptionsexperiment wurden Studierenden in Bern und Zürich, die sich als Mundartsprecher bezeichnen, zuerst die vier nicht bearbeiteten Beispiele einmal vorgespielt. Die Testpersonen mussten direkt nach jedem Beispiel in einer *forced choice* angeben, um welche Mundart es sich handelt. Dadurch wird überprüft, wie gut die Zuordnung zu den vier schweizerdeutschen Mundarten ist. Anschließend wurden die 16 delexikalisierten Beispiele in randomisierter Folge abgespielt und wiederum eine Zuordnung zu den vier Mundarten verlangt.

Die nicht bearbeiteten Beispiele werden trotz deren Kürze in gut 88 % der Fälle korrekt zugeordnet. Das Ergebnis liegt deutlich über dem Zufallsniveau von 25 %. Somit ist eine Erkennung der Dialekte als gegeben zu erachten. Die Aufgabe, die delexikalisierten Beispiele zu erkennen, erweist sich als ungleich schwieriger. Die richtige Zuordnungsrate liegt auch nur bei gut 32 % (7 % über dem Zufallsniveau), ist aber signifikant. Die beste Erkennungsquote erreichen

die Berner (39 %), dicht gefolgt von den Walliser Aufnahmen (37 %). Weniger gut erkannt werden die Zürcher Aufnahmen (32 %), und die Erkennungsquote der Bündner Aufnahmen liegt mit 19 % sogar unter dem Zufallsniveau. Die Verteilung innerhalb der Mundarten zeigt einige bemerkenswerte Resultate: So erkennen die Berner und Zürcher Versuchspersonen jeweils ihre eigene Mundart schlechter als die andere Mundart. Interessant ist auch die unterschiedliche Erkennungsquote der Alpinmundarten: Während die (westlichen) Berner Studierenden den (westlichen) Walliserdialekt signifikant besser erkennen als die (östlichen) Zürcher Studierenden, ist das Ergebnis beim (östlichen) Bündner Dialekt genau umgekehrt. Die beiden ‚exotischeren‘ Mundarten werden also von den direkten Nachbarn besser erkannt, was in Wahrnehmungsstudien als *proximity effekt* bezeichnet wird (Montgomery 2007: 151–159; Purschke 2011). Als weiterer Punkt ist anzumerken, dass Berner und Zürcher bzw. Berner und Walliser Beispiele kaum verwechselt wurden, andererseits Walliser und Zürcher öfters. Das deutet darauf hin, dass das Timing und dabei wohl besonders die Artikulationsgeschwindigkeit eine wichtige Rolle in der Perzeption von Sprache darstellt und evtl. wichtiger ist als die Intonation.

Somit kann festgehalten werden, dass die Prosodie der vier gewählten Schweizer Mundarten so klar unterschiedlich ist, dass eine Zuordnung allein auf der Basis der Prosodie schwierig, aber grundsätzlich möglich ist. Die eigene Varietät wird dabei eher schlechter erkannt als eine fremde, was daran liegt, dass sie als ‚normal‘ angesehen wird, während die Prosodie der anderen Mundarten auffällig erscheint. Gleichzeitig spielt die geographische und damit verbunden die kommunikative Nähe eine Rolle, da die näheren, aber von der eigenen Mundart unterschiedlichen Beispiele besser erkannt werden als die ferneren. Der Grad der Falschzuordnungen deutet zudem darauf hin, dass

schnelle und langsame Mundarten kaum verwechselt werden, so dass Timing und Artikulationsgeschwindigkeit im Bereich der Perzeption der Prosodie wohl eine größere Rolle spielen als die Intonation, welche ja zumindest teilweise auch vom Timing abhängt.

## 2.4 Erkenntnisgewinn

Die Studie hat gezeigt, dass auch innerhalb einer Dialektgruppe – dem Hoch- und Höchstalemannischen – prosodische Unterschiede vorhanden sind und dass sie auch von Laien erkannt werden. Die Erkenntnis aus dem Projekt liegt vor allem in der empirischen und quantifizierbaren Darstellung dieser Unterschiede und dem dazu erarbeiteten methodischen Inventar.

Die Analyse der Sprachdaten wie auch der Perzeptionstest zeigt den Sonderstatus des Walliserdeutschen auch in prosodischer Hinsicht. Besonders auffällig ist die hohe Sprech- und Artikulationsgeschwindigkeit sowie die relativ geringe Markierung von Phrasengrenzen mittels Dehnungen. Sowohl im Timing als auch in der Intonation zeigt sich ein geringer Einfluss von Wortklasse und Wortakzent (*stress*), dafür finden sich besonders viele Akzente, die mit Schwa-Silben in Verbindung gesetzt werden, womit das Walliserdeutsche eine stärkere Tendenz zur Silbensprachlichkeit aufweist als die anderen untersuchten Mundarten. Die Walliser Prosodie ist generell sehr variabel und schlecht voraussagbar. Im Weiteren ist das Berndeutsche auffällig, das in vielem ein Gegenpol zum Walliserdeutschen darstellt: Es ist die am langsamsten realisierte Mundart. Phrasengrenzen sind deutlich im Timing markiert, in der Intonation zeigt das Berndeutsche auffällig späte Akzentkommandos, die eine Anbindung an die lexikalische Kategorie *stress* aufweisen. Diese beiden Mundarten zeigen in der Erkennung delexikalisierter

Prosodie die höchsten Erkennungsquoten. Das Zürichdeutsche ist auch eher schnell gesprochen, Phrasen sind im Timing deutlich markiert, *stress* und Wortklasse sind im Timing und in der Intonation bedeutsam, Akzentkommandos sind nicht so spät wie im Berndeutschen. Es zeigt damit am ehesten Merkmale, die mit intonatorischen Merkmalen der Standardsprache verbunden werden und damit vermutlich als weniger auffällig gelten können (vgl. dazu Ulbrich 2005, die die Intonation von bundesdeutschen, österreichischen und Schweizer Nachrichtensprechern untersucht, womit auch eine Schweizer Variante des Standards beschrieben wird, oder Peters 2006, der für sieben bundesdeutsche Städte jeweils standardnähere und dialektnähere Sprecher unterscheidet). Das Bündnerdeutsche weist im Timing und in der Intonation die beste Voraussagbarkeit und damit die einfachste Struktur aus. In der Intonation sind die hohen, flachen  $F_0$ -Konturen und der hohe Anteil fallender Akzente auffällig. Mit dem Walliserdeutschen verbindet es der fehlende Einfluss von Wortklasse und *stress* auf die Prosodie sowie die eher geringe temporale Markierung der Phrasengrenzen, mit dem Zürichdeutschen die eher hohen Phrasenkommandos und die Ähnlichkeit der Verteilung der intrinsischen Lautdauer. Eine sprachgeographische Strukturierung im Sinne einer Nord / Süd- bzw. Ost / West-Gliederung konnte mit den bestehenden Daten eingeschränkt erfasst werden. Als Nord / Süd-Unterschied zeigt sich einerseits die Bedeutung des lexikalischen Akzents (*stress*) und der Wortklasse für die Prosodie nördlicher Mundarten, andererseits die deutlichere Tendenz zur Dehnung als Mittel zur Markierung prosodischer Phrasengrenzen in den nördlichen Mundarten. Als West / Ost-Gegensatz ist insbesondere das unterschiedliche Verhältnis der intrinsischen Lautdauer und der stärkeren intonatorischen Bedeutung globaler Aspekte in den östlichen Mundarten zu nennen.



In methodischer Hinsicht wurde deutlich, dass sowohl Timing- wie auch Intonationsaspekte für eine prosodische Charakterisierung notwendig sind und dass sich die beiden Aspekte nur teilweise decken. Gleichzeitig wurde deutlich, dass die Interaktionen verschiedener Faktoren sehr komplex sind, so dass eindeutige Resultate mit der Analyse von Spontansprache wohl nicht zu erreichen sind. Die Verbreiterung der Datenbasis gegenüber dem Vorgängerprojekt verdeutlicht die dialektalen Unterschiede; deren Überlagerung durch individuelle Charakteristika sind aber immer noch deutlich, was sich in der häufig besseren individuellen Modellierung gegenüber der dialektalen Modellierung äußert. Der gewählte quantitative Ansatz kann bisherige impressionistische Beschreibungen im Bereich der Intonation empirisch unterstützen, ermöglicht aber erstmals systematische und akustische Vergleiche. Im Bereich des Timings zeigen sich mit diesem Ansatz erstmals überhaupt systematische Darstellungen, da Vergleiche der Dialekte ohne empirische Datenbasis kaum möglich waren und in der Literatur außer mit allgemeinen Hinweisen auf die Sprechgeschwindigkeit kaum vorgenommen wurden.

### 3 Ausblick

Die Datenaufbereitung für eine quantitative Analyse ist überaus aufwändig und hat fast die Hälfte der Arbeitszeit im Projekt eingenommen, so dass die Zeit für die Analyse eher knapp war. Weiterführende Analysen mit den bestehenden Daten sind immer noch möglich. Eine baldige sprachgeographische Vertiefung der Erkenntnisse wird wohl nicht so schnell zu erreichen sein, wenn nicht neue automatisierte Methoden zur Segmentation von Aufnahmen entwickelt werden. Doch auch dann bleibt für die Annotation der Daten noch ein beträchtlicher

Aufwand zu betreiben, weil dafür – möglichst muttersprachliches – linguistisches Wissen notwendig ist. Gleichwohl ist ein quantitativer Ansatz erfolgversprechend, weil impressionistische Beschreibungen regionaler Intonation so auf eine empirisch sichere Basis gestellt werden können. Im Bereich des Timings werden auf diese Weise überhaupt erst belegbare Resultate möglich, da Timingaspekte außer mit Hinweisen zur allgemeinen Sprechgeschwindigkeit impressionistisch nur wenig beschrieben worden sind. Zudem ist das Timing regionaler Prosodie auch außerhalb des deutschen Sprachraums bislang kaum erforscht worden, was als generelle Forschungslücke festgehalten werden kann. Die Detailanalysen haben auch gezeigt, dass für das Timing eine Segmentation auf Lautebene notwendig ist, weil sich segmentale Aspekte auf höherer Ebene (Silbe und Phrase) auswirken. Im Bereich der Intonation kann das gewählte Command-Response-Modell gut zur Beschreibung genutzt werden, da die Datenaufbereitung noch keine linguistische Information und Interpretation verwendet, wie sie für eine Analyse mittels Tonsequenzmodellen wie ToBI schon viel früher im Analyseprozess notwendig ist. Offen geblieben ist in der Analyse die Intensität, die generell phonetisch noch wenig untersucht ist, weil die Daten häufig – auch in diesem Projekt – dafür zu wenig adäquat erfasst worden sind.

Die Modellierung im Bereich des Timings und noch viel mehr im Bereich der Intonation hat deutlich gemacht, dass viele Faktoren, die die spontansprachliche Prosodie mitbestimmen, noch gar nicht erfasst worden sind. Um diese Variation – vermutlich v. a. im Bereich Informations- und Gesprächsstrukturierung – einzuschränken oder zu kontrollieren, sind entsprechende Erhebungs- oder Auswahlmethoden zu entwickeln, welche die Spontansprachlichkeit allerdings möglichst wenig beeinflussen sollen.

## 4 Literatur

- Boersmaa, Paul & David Weenink (2014): *Praat: doing phonetics by computer [Computer program]. Version 5.3.86*. online. <http://www.praat.org> (1.10.2014)
- Christen, Helen (1998): *Dialekt im Alltag. Eine empirische Untersuchung zur lokalen Komponente heutiger schweizerdeutscher Varietäten*. Tübingen: Niemeyer.
- Fiehler, Reinhard, Birgit Barden, Mechtild Elstermann & Barbara Kraft (2004): *Eigenschaften gesprochener Sprache*. Tübingen: Gunter Narr.
- Fitzpatrick-Cole, Jennifer (1999): The alpine intonation of Bern Swiss German. In John J. Ohala (ed.): *Proceedings of the XIVth International Congress of Phonetic Sciences (ICPhS)*, 941–944. San Francisco.
- Fujisaki, Hiroya & Keikichi Hirose (1982): Modelling the dynamic characteristics of voice fundamental frequency with applications to analysis and synthesis of intonation. In *Preprints of the Working Group on Intonation, 13th International Congress of Linguists*. Tokyo: 57–70.
- Gilles, Peter (2005): *Regionale Prosodie im Deutschen. Variabilität der Intonation von Abschluss und Weiterweisung*. Berlin, New York: De Gruyter.
- Gilles, Peter & Renate Schrambke (2000): Divergenz in den Intonationssystemen rechts und links des Rheins. Die Sprachgrenze zwischen Breisach (Baden) und Neuf-Breisach (Elsass). In Edith Funk, Werner König & Manfred Renn (Hrsg.): *Bausteine zur Sprachgeschichte. Referate der 13. Arbeitstagung zur alemannischen Dialektologie in Augsburg 29.9. - 3.10.1999*, 87–98. Heidelberg: Winter.
- Grosjean, François & Michael Collins (1979): Breathing, pausing, and reading. *Phonetica* 36: 98–114.
- Häsler, Katrin, Ingrid Hove & Beat Siebenhaar (2005): Die Prosodie des Schweizerdeutschen - Erkenntnisse aus der sprachsynthetischen Modellierung von Dialekten. *Linguistik online* 24, 3/05: 187–224. <https://bop.unibe.ch/linguistik-online/article/view/643> (1.10.2014)

- Hotzenköcherle, Rudolf (1984): *Die Sprachlandschaften der Schweiz*. Aarau u. a.: Sauerländer.
- Leemann, Adrian (2012): *Swiss German Intonation Patterns*. Amsterdam: Benjamins.
- Leemann, Adrian & Beat Siebenhaar (2008a): Swiss Alpine and Midland Intonation. In Plínio A. Barbosa, Sandra Madureira & César Reis (eds.): *Proceedings of the Speech Prosody 2008 Conference*, 289–292. Campinas, Brazil: Editora RG/CNPq.
- Leemann, Adrian & Beat Siebenhaar (2008b): Perception of Dialectal Prosody. In *Proceedings of Interspeech 2008, Brisbane, Australia, 22.-26.9.2008*, 524–527. Brisbane.
- Leemann, Adrian & Beat Siebenhaar (2010): Statistical Modeling of F<sub>0</sub> and Timing of Swiss German Dialects. In *Proceedings of Speech Prosody 2010*, 1–4. Chicago. <http://speechprosody2010.illinois.edu/papers/100180.pdf> (1.10.2014)
- Mixdorff, Hansjörg (2009): Program for Estimating Fujisaki-Parameters. Unpublished manual. <http://public.tfh-berlin.de/~mixdorff/thesis/fujisaki.html> (1.10.2014).
- Moulton, William G. (1986): Sandhi in Swiss German dialects. In Henning Andersen (ed.): *Sandhi Phenomena in the Languages of Europe*, 385–392. Berlin: De Gruyter.
- Montgomery, Christopher (2007): *Northern English dialects: A perceptual approach*. Sheffield: University of Sheffield. <http://core.kmi.open.ac.uk/download/pdf/43192.pdf> (1.10.2014)
- Nübling, Damaris & Renate Schrambke (2004): Silben- versus akzentsprachliche Züge in germanischen Sprachen und im Alemannischen. In Elvira Glaser, Peter Ott & Ruedi Schwarzenbach (Hrsg.): *Alemannisch im Sprachvergleich. Beiträge zur 14. Arbeitstagung für alemannische Dialektologie in Männedorf (Zürich) vom 16. - 18.9.2002*, 281–320. Stuttgart: Franz Steiner.
- Peters, Jörg (2006): *Intonation deutscher Regionalsprachen*. Berlin, New York: De Gruyter.
- Purschke, Christoph (2011): Regional linguistic knowledge and perception on the conceptualization of Hessian. *Dialectologia* Special Issue II: 91–118. <http://www.raco.cat/index.php/Dialectologia/article/view/245346/328592> (1.10.2014)

- Schnidrig, Kurt (1986): *Das Dusseln. Ein Subsidiärdialekt im Deutschwallis*. Freiburg i. Ue.: Univeristätsverlag.
- Ris, Roland (1992): Innerethik der deutschen Schweiz. In Paul Hugger (Hrsg.): *Handbuch der schweizerischen Volkskultur, Band II*, 749–766. Zürich: Offizin.
- Siebenhaar, Beat (2012): Wie ‚singen‘ Dialektsprecher? – Zur Erforschung der Sprechmelodien in den Dialekten. In Rainer Hünecke & Karlheinz Jakob (Hrsg.): *Die obersächsische Sprachlandschaft in Geschichte und Gegenwart*, 265–288. Heidelberg: Universitätsverlag Winter.
- Siebenhaar, Beat (2014): Phonological and phonetic considerations for a classification of Swiss German dialect as a word language or syllable language. In Renata Szczepaniak, Renata & Javier Caro Reina (eds.): *Phonological Typology of Syllable and Word Languages in Theory and Practice*, 327–345. Berlin: De Gruyter.
- Siebenhaar, Beat, Brigitte Zellner Keller & Eric Keller (2001): Phonetic and Timing Considerations in a Swiss High German TTS System. In Eric Keller, Gérard Bailly, Alex Monaghan, Jacques Terken & Mark Huckvale (eds.): *Improvements in Speech Synthesis*, 165–175. Chichester: Wiley.
- Siebenhaar, Beat & Adrian Leemann (2012): Methodological reflections on the phonetic–phonological continuum, illustrated on the prosody of Swiss German dialects. In Bernhard Wälchli, Andrea Ender & Adrian Leemann (eds.): *Methods in Contemporary Linguistics*, 21–44. Berlin: De Gruyter.
- Stalder, Franz Joseph (1819): *Die Landessprachen der Schweiz oder Schweizerische Dialektologie, mit kritischen Sprachbemerkungen beleuchtet. Nebst der Gleichnisrede vom verlorenen Sohne in allen Schweizermundarten*. Aarau: Sauerländer.
- Szczepaniak, Renata (2007): *Der phonologisch-typologische Wandel des Deutschen von einer Silben- zu einer Wortsprache*. Berlin, New York: De Gruyter.
- Ulbrich, Christiane (2005): *Phonetische Untersuchungen zur Prosodie der Standardvarietäten des Deutschen in der Bundesrepublik Deutschland, in der Schweiz und in Österreich*. Frankfurt am Main: Lang.

Werlen, Iwar (2005): Mundarten und Identität. In Paolo Barblan & Arnold Koller (Hrsg.): *Dialekt in der (Deutsch)Schweiz – Zwischen lokaler Identität und nationaler Kohäsion*, 26–32. Lenzburg: Forum Helveticum.

Werlen, Iwar, Barbara Buri, Marc Matter & Johanna Ziberi (2002): *Projekt Üsserschwyz: Dialektanpassung und Dialektloyalität von Oberwalliser Migranten*. Bern: Institut für Sprachwissenschaft der Universität Bern.

Wipf, Elisa (1910): *Die Mundart von Visperterminen im Wallis*. Frauenfeld: Huber.

Zimmermann, Gerhard (1998): Die ‚singende‘ Sprechmelodie im Deutschen. Der metaphorische Gebrauch des Verbums ‚singen‘ vor dem Hintergrund sprachwissenschaftlicher Befunde. *Zeitschrift für Germanistische Linguistik* 26: 1–16.