

## Berner und Zürcher Prosodie – Ansätze zu einem Vergleich

Hier werden vorläufige, zu ergänzende, vielleicht sogar zu widerrufende Ergebnisse präsentiert, die sich aus einem Nationalfonds-Projekt zur Erforschung der schweizerdeutschen Prosodie mittels sprachsynthetischer Modellierung ergeben. Für viele wertvolle Hinweise danke ich den ProjektmitarbeiterInnen Eric Keller und Martin Forst, Ingrid Hove, Katrin Häslar sowie Brigitte Zellner Keller.

### 1 Dialektologie und Prosodie

Die Dialektologie hat sich lange auf die Realisierung von Einzellauten konzentriert. Die Prosodie blieb weit gehend unerforscht, obwohl sie von Linguisten und Laien häufig als auffällig genannt wird<sup>1</sup>. So ist für R. Hotzenkocherle (1984, 1995) die "Eigenart einer schwer zu beschreibenden, aber höchst ohrenfälligen, charakteristischen Intonation" ein konstituierendes Merkmal für das Berndeutsche<sup>2</sup>.

Für die Beschäftigung mit Prosodie fehlten lange die nötigen Instrumente. Erst mit erschwinglichen Computerprogrammen zur apparativen Phonetik findet Prosodieforschung eine weitere Verbreitung. Nun werden zuerst Standardsprachen und standardnahe Varietäten, in jüngster Zeit auch Dialekte erforscht. In diskursanalytischen Ansätzen werden einerseits bestimmte Intonationskonturen gesucht und deren Funktion untersucht, andererseits kommunikative Funktionen und deren intonatorische Umsetzung. Diesen Weg verfolgt das DFG-Projekt "Untersuchungen zur Struktur und Funktion regional-spezifischer Intonationsverläufe im Deutschen"<sup>3</sup>. Für die deutschen Dialekte haben P. Gilles und R. Schrambke (2000) die 'Divergenz in den Intonationssystemen links und rechts des Rheins' untersucht. In den beiden niederalemannischen Mundarten finden sich unterschiedliche Intonationsmuster für dieselben kommunikativen Funktionen. J. Fitzpatrick-Cole (1999) vergleicht Berner Intonationskonturen mit denjenigen norddeutsch geprägter Standardsprache. Dabei sieht sie einen wesentlichen Unterschied im unterschiedlichen default accent: betonter Hochton mit folgendem unbetontem Tiefton (H\*+L) für die Standardsprache und betonter Tiefton mit folgendem unbetontem Hochton (L\*+H) für das Berndeutsche. In diesem Band findet sich mit der Darstellung der 'Silbensprachlichen Strukturen im Schweizerdeutschen' (D. Nubling / R. Schrambke, in diesem Band) ein weiterer Aspekt. Prosodie rückt also auch ins Interesse der dialektologischen Forschung, wie das D. Hirst und A. Di Cristo (1998, 43) in ihrem 'survey of intonation systems' fordern und wie es in diesem Sammelband vereinzelt angesprochen wird. So werden für das Deutsche einzelne Phänomene erwähnt, die Mundarten unterscheiden (D. Gibbon 1998).

<sup>1</sup> Zur Auffälligkeit regionaler Prosodie vgl. G. Zimmermann (1998).

<sup>2</sup> Dass Hotzenkocherle die Intonation für das Berndeutsche als "ohrenfällig" beurteilt, ist nicht erstaunlich, wenn man bedenkt, dass er selbst einen ostschweizer Dialekthintergrund hat, gegenüber dem das fremde Intonationssystem eben auffällt.

<sup>3</sup> Siehe dazu die im Literaturverzeichnis gegebenen Publikationen von P. Auer, P. Gilles, J. Peters, M. Selting.

Wie in der gesamten Prosodieforschung stehen auch in der deutschen Variationslinguistik Aspekte der Intonation in Zentrum, während Phrasierung und Timing eher peripher behandelt werden. Diese Schwerpunktsetzung ist jedoch aus mehreren Gründen zu hinterfragen. Denn: grundsätzlich ist jede Tätigkeit, jede Veränderung in ein zeitliches Raster eingebunden. So ist die Zeit auch in der Sprache als primärer Faktor zu beurteilen, auf den die Veränderungen im Bereich der Grundfrequenz, der Intensität und des Spektrums zu beziehen sind. Veränderungen des Sprachsignals sind teilweise durch physiologische Gegebenheiten der Sprachproduktion eingeschränkt, welche eine zu schnelle Modifizierung nicht erlauben. Zu schnelle Modifikationen führen zu Assimilationen, d.h. zu unfertigen artikulatorischen Gestus. Zudem ist die Zeitachse interindividuell stabiler als die anderen Aspekte: E. Keller (1994)<sup>4</sup> hat im Vergleich von zwölf Realisierungen desselben Satzes gezeigt, dass die Silbendauer in gelesenen Texten zwischen verschiedenen Sprechern viel stärker korreliert als der Grundfrequenzverlauf. Informelle Tests mit Segmentdauermodifikationen von französischen Vokalen haben ergeben, dass systematische Veränderungen der Segmentdauer um 2% wahrgenommen werden, während Fehler im Intonationsverlauf weniger auffallen. Diese Erkenntnis wird gestützt durch Perzeptionstests mit verschiedenen deutschen Sprachsynthesystemen (H. Mixdorff 2002), in denen das Modell mit der besten zeitlichen Übereinstimmung zu natürlichen Daten am besten bewertet wird. B. Zellner Keller (i.pr.) folgert aus ihren vielen Arbeiten zur zeitlichen Organisation in der Sprache (B. Zellner 1994, 1998, B. Zellner Keller 2002a, 2002b), dass "rhythm should be seen as a key prosodic tool for signaling an overarching semantic organisation." Diesem Ansatz folgt auch das Projekt zur Modellierung der Mundartprosodie.

### 3      Erarbeitung von Grundlagen zur Erforschung schweizerdeutscher Prosodie mittels sprachsynthetischer Modellierung<sup>5</sup>

Das Projekt setzt andere Schwerpunkte als die meisten deskriptiven und diskursanalytischen Zugänge. Wir versuchen die Analysen zu Modellen weiterzuentwickeln und die verschiedenen Modelle in eine Synthese zusammenzuführen. Das bedingt, dass jeder Teil der Sprache abgedeckt wird, denn eine Sprachsynthese funktioniert nicht, wenn sie eine Sprache nicht vollständig abdeckt. Es nützt nichts, – beispielsweise – das beste Modell für die Intonation in Weiterweisungen (cf. P. Gilles 2001) zu haben, wenn kein Modell für den Satzanfang, für das Satzende existiert, wenn kein Modell für die Phrasierung besteht, wenn wir nicht wissen, wie lange die einzelnen Segmente sind, wenn wir nicht wissen, durch welche Faktoren sie gedehnt oder gekürzt werden. Die Sprachsynthese verlangt also nach einer möglichst vollständigen Beschreibung der Sprache und deren Umsetzung in Algorithmen. Somit muss der Schwerpunkt der Analyse für eine Synthese vorerst in globalen Aspekten liegen, die allmählich verfeinert werden kann. Es geht also zuerst darum zu wissen, wie ein unmarkierter Satz prosodisch aussieht; es geht darum, ein Modell zu haben für die Positionen, wo natürliche Sprecher pausieren, ein Modell für die Dauer der einzelnen Phoneme; ein Modell für die 'normale' Intonation, mit der ein Sprecher einen Satz realisiert.

---

<sup>4</sup> auf der Basis von Daten aus Caelen-Haumont (1991)

<sup>5</sup> Titel des Nationalfondsprojekts unter der Leitung von Eric Keller und den Mitarbeitern Beat Siebenhaar, Martin Forst (bis Sept. 2002), Ingird Hove und Katrin Häslar (beide seit Okt. 2002).

Die verschiedenen Modelle werden dann in die Sprachsynthese integriert, die Texte vorsprechen kann. Mit diesen synthetischen Äußerungen wird es möglich, verschiedene Modelle auditiv zu überprüfen und anhand einer solchen Fehleranalyse die Modelle zu verfeinern<sup>6</sup>. Die Synthese ist also eine Methode, prosodische Modelle hörbar zu machen. Das Funktionieren der Sprachsynthese soll in Grundzügen dargestellt werden, um die nachfolgend präsentierten Ergebnisse zu motivieren.

## 4 Sprachsynthese

Was wird für eine Sprachsynthese benötigt? Als erstes einen vorzulesenden Text. Für die Mundartsynthese verlangen wir eine möglichst lautgetreue Schreibung, so kann der Aufwand für die Programmierung der Graphem-Phonem-Übersetzung klein gehalten werden, die im Rahmen eines Projekts zur Prosodiesteuerung wenig interessiert.

Als zweites brauchen wir eine Stimme. Ein Großteil der aktuellen Sprachsynthesen verwendet zur Signalgeneration eine so genannte Diphonkonkatenation, d.h. Diphone, Kombinationen von zwei Phonemen, werden aneinandergehängt und in Dauer und Grundfrequenz modifiziert. Um zu Diphonen zu kommen, wird ein menschlicher Sprecher aufgenommen, der Sätze mit allen möglichen Phonemkombinationen spricht. Diese Diphone werden aus dem Signal ausgeschnitten und in eine Datenbank abgelegt, auf die das Sprachsyntheseprogramm zurückgreift. Die einzelnen Diphone werden für das Sprachsignal zusammengehängt und so modifiziert, dass ihre Dauer und Grundfrequenz den errechneten Werten entspricht.

Zwischen Eingabe und Signalgeneration ist der Teil, der hier besonders interessiert: Das prosodische Modul, das Phrasengrenzen bestimmt, die Dauer von Segmenten und die Grundfrequenz. Es errechnet also aus einer abstrakten Folge von Zeichen, welche Diphone ausgewählt werden müssen, wie lange die einzelnen Laute sein müssen und mit welcher Grundfrequenz sie abzuspielen sind. In ihren Grundzügen kann die im Folgenden dargestellte Methode zur Errechnung dieser Werte, wie sie in unserem Sprachsyntheseprogramm für die französische und deutsche Standardsprache, LAIPTTS<sup>7</sup>, zur Anwendung gelangt, für die Mundartsynthese übernommen werden. Es wäre jedoch uberaus erstaunlich, wenn keine varietätenspezifischen Modifikationen nötig wurden.

### 4.1 Phrasierung

Das von Keller und Zellner Keller für die französische Lesesprache entwickelte, mehr von psycholinguistischen als von syntaktischen Prinzipien geleitete Phrasierungsmodell geht davon aus, dass Sprecher die zu vermittelnden Inhalte in kleinere Portionen aufteilen. Diese Phrasen sind selten länger als 12 Silben. Zudem neigen Sprecher zu einer rhythmisch ausgewogenen Phrasierung, sie tendieren dazu, verschiedene Phrasen im Satz mehr oder weniger gleich lang zu machen. Diese beiden Prinzipien teilen einen Satz an 'Sollbruchstellen' auf, die nur teilweise syntaktischen Grenzen entsprechen. Ansätze zu einer Übertragung auf die gesprochene Sprache finden sich weiter unten.

---

<sup>6</sup> Die methodischen Grundlagen für die Synthese als Forschungsinstrument finden sich in B. Siebenhaar (i.Dr.) und B. Siebenhaar, M. Forst, E. Keller (i. pr.)

<sup>7</sup> Die französische Synthese wurde von Eric Keller und Brigitte Zellner Keller entwickelt. Die da geschaffenen Modelle wurden von Beat Siebenhaar und Martin Forst für eine Synthese der deutschen Lesesprache übertragen.

## 4.2 Timing

Timing ist eng verzahnt mit der Phrasierung. Denn neben der intrinsischen Dauer der einzelnen Segmente und der Beeinflussung durch die benachbarten Segmente zeigen die Phrasengrenzen einen wesentlichen Einfluss auf die Silben- und damit die Segmentdauer. Ebenfalls sind der Wortakzent und der grammatische Status eines Wortes von Bedeutung. Das Gewicht der einzelnen Parameter wird mittels einer statistischen Analyse gewonnen. Erste Ergebnisse dieser Analyse für die Mundartsynthese werden anschließend vorgestellt.

## 4.3 Intonation

Nachdem die Dauer der einzelnen Segmente berechnet ist, wird eine Intonation errechnet. Dabei orientieren wir uns am Fujisaki-Modell, das von längeren Phrasenkomponenten und dazuaddierten kürzeren Akzentkommandos ausgeht (zuerst präsentiert von H. Fujisaki und K. Hirose 1982, für eine Adaptation für das Deutsche siehe H. Mixdorff 2002).

## 5 Daten und Datenaufbereitung

Prosodie ist gekennzeichnet durch eine große Variabilität, deren Ursachen erst zu einem kleinen Teil geklärt sind. Klar ist, dass Variation innerhalb derselben Varietät interindividuell und intraindividuell begründet ist. Für analytische Arbeiten zur Prosodie werden häufig mehrere Personen untersucht, durch die Auswahl der 'typischen' Fälle wird die Variation je nach dem reduziert oder erklärt. Für die Sprachsynthese dagegen wird meist nur eine Gewährsperson untersucht, so wird interindividuelle Variation ausgeschlossen. Es ist der klassische Weg der Gewährsleutedialektologie. Für das vorliegende Projekt zur Prosodie des Berndeutschen und des Zürichdeutschen wird dieser zweite Weg gewählt.

Wie für jede Arbeit, die sich auf Gewährsleute stützt, ist deren Auswahl einer der heiklen Punkte. Da die Berner und Zürcher Prosodie nicht untersucht ist, kann die Auswahl der Gewährsleute nur schwer begründet werden. Wir haben deshalb zwei Sprecher ausgewählt, deren Mundart auf segmentaler Ebene deutlich einem der beiden Mundarträume zuzuordnen sind. Sie entsprechen zudem den traditionellen dialektologischen Auswahlkriterien in Bezug auf die Ortsansässigkeit. Von beiden Sprechern wurden Aufnahmen in ähnlichen Interviewsituationen analysiert. Stilistisch unterscheiden sich die beiden Sprecher insofern, als der Zürcher etwas bedächtiger ist, jedoch weniger Selbstkorrekturen vornimmt, während der Berner weniger sorgfältig formuliert und häufiger Selbstkorrekturen aufweist. Die Aufnahmen für den Berner Sprecher umfassen ein Interview von 20 Minuten, für den Zürcher Sprecher sind es zwei Interviews von zusammen 50 Minuten. Die Aufnahmen wurden mit Hilfe eines Aligners vorsegmentiert und manuell korrigiert<sup>8</sup>. Daraus ergaben sich für den Zürcher Sprecher 16'436 Segmente, für den Berner Sprecher sind dies 8'462. Aus diesen Daten wird nun für jede Mundart ein Prosodiemodell erstellt. Wir stehen im Prozess der Analyse, in welcher die relevanten Faktoren für die Modellierung herausgearbeitet werden.

---

<sup>8</sup> Vielen Dank an Martin Forst, der einen großen Teil dieser Arbeit geleistet hat.

## 6 Analyse

### 6.1 Phrasierung

Phrasierung, die Strukturierung von Äußerungen in kleinere Einheiten, bestimmt die Wahrnehmung der Sprachmelodie und des Sprachrhythmus wesentlich. Die Strukturierung in spontanen Äußerungen in der Mundart wird im Folgenden mit den für die Lesesprache entwickelten Prinzipien verglichen.

Für die französische Sprachsynthese hat sich gezeigt, dass eine auf psycholinguistischen Prinzipien beruhende rudimentäre Syntax die Phrasierung gut modellieren kann. Die Grundlagen für die Phrasierung nach so genannten Performanzstrukturen finden sich in Arbeiten von F. Grosjean und Mitarbeitern (u.a. F. Grosjean / M. Collins 1978, J. Gee / F. Grosjean 1983, P. Monnin / F. Grosjean 1993). Weiterführende empirische Untersuchungen führten zur Ausarbeitung eines Algorithmus für die französische Lesesprache (E. Keller et al. 1993; E. Keller / B. Zellner 1996; B. Zellner 1994, 1996, 1998, B. Zellner Keller 2002b), der in die Sprachsynthese LAIPTTS\_F eingebaut ist.

Das Modell geht vom Satz aus. Auf der Unterscheidung von lexikalischen und grammatischen Wörtern wird der Zusammenhang der Wörter festgestellt. Dieser Zusammenhang ist stark zwischen grammatischen Wörtern und nachfolgenden lexikalischen Wörtern, während der Zusammenhang da schwach ist, wo ein grammatisches Wort einem lexikalischen Wort folgt; Ausnahmen wie rechts stehende Pronomen werden zusätzlich beachtet. Sollbruchstellen stehen also vor einzelnen oder mehreren Präpositionen, Konjunktionen, Artikeln und Pronomen. Anschließend wird der Satz in Phrasen aufgeteilt, wobei die erste Strukturierung sich an der Interpunktion orientiert. Falls Phrasen übrig bleiben, die länger als zwölf Silben sind, werden sie an der Sollbruchstelle, die sich am nächsten der Mitte findet, aufgeteilt. So wird ein regelmäßiger Rhythmus erreicht.

Für die Synthese einer deutschen Lesesprache schweizerischer Prägung, LAIPTTS\_D, konnte dieses Verfahren mit geringen Modifikationen übernommen werden. Die wesentlichste Änderung betrifft eine zusätzliche Sollbruchstelle vor dem flektierten Verb. Eine theoretische psycholinguistische Erklärung ist noch ausstehend. Es ist aber zu vermuten, dass die häufig komplexe morphologische und syntaktische Struktur im Zusammenhang mit dem deutschen Verb zu Pausen und Häsitationen führen, die beim französischen Verb nicht beobachtet werden können. Die psycholinguistische Forschung geht davon aus, dass Pausen an Stellen planerischer Unsicherheit vorkommen und somit Hinweise auf eine erhöhte kognitive Verarbeitungsdichte geben können.

Die Synthese spontan gesprochener Mundart verlangt weitergehende Modifikationen des Algorithmus. Mit dem obigen Verfahren werden nur 57% der Pausen in den Interviews richtig erfasst. Das ist viel weniger als die 90%, die sich für die standardsprachliche Lesesprache ergaben.

Für die Analyse wurden die Phrasengrenzen perzeptiv beurteilt. Dabei wurde unterschieden in Phrasengrenzen, a) die Sätze oder Turns syntaktisch und intonatorisch abschließen, b) die auf Satzebene motiviert sind und Weiterführung markieren, c) die auf Satzebene nicht motiviert sind und Weiterführung markieren, d) die als inadäquate Überlegungs- und Häsitationspausen wahrgenommen werden. Alle Phrasengrenzen können mit gefüllten oder stillen Pausen

kombiniert werden. Nicht berücksichtigt wurden Interpretationen auf einer höheren Ebene, z.B. Sprechakte, wie sie von W. N. Campbell (1997) vorgeschlagen werden.

Der Vergleich der Phrasenlänge zeigt Gemeinsamkeiten und Differenzen zwischen dem Zurcher und dem Berner Sprecher. Aus Abbildung 1 wird deutlich, dass bei beiden Sprechern kaum Phrasen vorkommen, die länger als 12 Silben sind. Die Ergebnisse zur maximalen Phrasenlänge können also von der gelesenen auf die gesprochene Sprache übertragen werden.

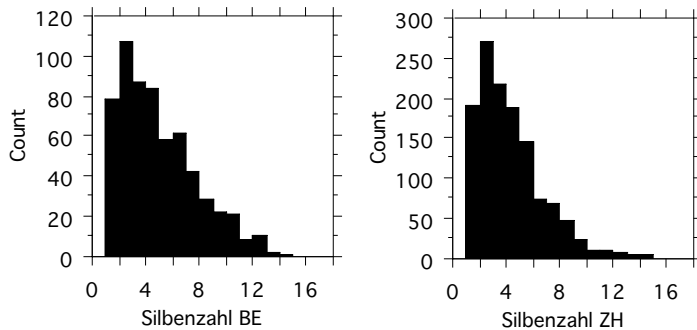


Abbildung 1: Histogramm der Phrasenlängen des Berner (links) und Zurcher (rechts) Sprechers

Unterschiede zwischen beiden Sprechern zeigen sich in der Verteilung. Der Modalwert liegt bei beiden Sprechern bei 2 Silben/Phrase, das arithmetische Mittel unterscheidet sich aber signifikant. Der Berner Sprecher hat einen Mittelwert von 4.5 Silben/Phrase, beim Zurcher Sprecher liegt er bei 3.8. Da unterscheiden sich die spontansprachlichen Daten von der gelesenen Standardsprache, wo mit einem Modalwert bei 8 Silben/Phrase und einem Mittelwert bei 7.6 Silben/Phrase annähernd eine Normalverteilung erreicht ist.

Gemäß dem oben dargestellten Phrasierungskonzept stehen Phrasengrenzen zwischen einem lexikalischen und einem folgenden grammatischen Wort, vor einem flektierten Verb oder in längeren Phrasen zwischen lexikalischen Wörtern. In den Daten beider Sprecher entsprechen 43% der Phrasen nicht diesem Muster.

Die beiden Sprecher zeigen eine unterschiedliche Verteilung der verschiedenen Phrasengrenzen (Tabelle 1): Während rund 38% aller Phrasengrenzen beim Berner Sprecher als Häsitationen gewertet wurden, sind dies beim Zurcher Sprecher nur 26%. Der Berner Sprecher zeigt mit einem Anteil von 26% fast doppelt so viele satzabschließende Phrasengrenzen als der Zurcher Sprecher, der nur 12% solcher Grenzen aufweist, aber an syntaktisch motivierten Grenzen häufiger intonatorisch weiterweisend fortfährt. Diese Unterschiede sind vermutlich individueller und nicht dialektaler Natur. Damit deuten sie auf ein Hauptproblem der Analyse hin, nämlich die Definition, welche Variation als individuell und welche als dialektal zu werten ist.

	BE n	BE%	ZH n	ZH%
a) abschließend	158	25.94	153	12.09
b) weiterführend, syntaktisch motiviert	68	11.17	232	18.34
c) weiterführend, nicht syntaktisch motiviert	154	25.29	555	43.87
d) Häsitationen	229	37.60	325	25.69
Total	609	100.00	1265	100.00

Tabelle 1: Frequenzverteilung der Phrasengrenzen

Die folgenden Abbildungen zeigen den Zusammenhang von Phrasenlänge und -grenze. Dabei wird unterschieden, ob die Phrasengrenze syntaktisch bedingt ist (Klassen a und b), ob sie als

angemessen (c) oder als unangemessen (d) beurteilt wurde. Für die folgende Beurteilung ist immer mit zu berücksichtigen, dass der Zürcher Sprecher generell kürzere Phrasen aufweist. Abbildung 2 zeigt die Phrasenlängen vor und nach Phrasengrenzen. Vor Phrasengrenzen weisen beide Sprecher die längsten Phrasen auf. Die kürzesten Phrasen finden sich vor Häsitations- und Überlegungsgrenzen. Dazwischen liegen die Grenzen, die nicht syntaktisch bedingt sind, aber als angemessen beurteilt wurden. Diese sind bei beiden Sprechern signifikant unterschieden zu den anderen Positionen. Die Unterschiede sind jedoch verschieden groß. Während beim Berner Sprecher Phrasen vor angemessenen und unangemessenen Grenzen ähnlicher sind, verhalten sich beim Zürcher Sprecher Phrasen vor angemessenen Grenzen ähnlicher wie diejenigen vor grammatischen Grenzen. Nach Phrasengrenzen zeigen beide Sprecher unterschiedliche Phrasenlängen. Die kürzesten Phrasen stehen bei beiden Sprechern nach unangemessen beurteilten Pausen. Von diesen unterscheiden beim Berner Sprecher diejenigen nach grammatisch begründeten Grenzen signifikant, beim Zürcher Sprecher sind es diejenigen nach richtig beurteilten, nicht syntaktisch bedingten Grenzen.

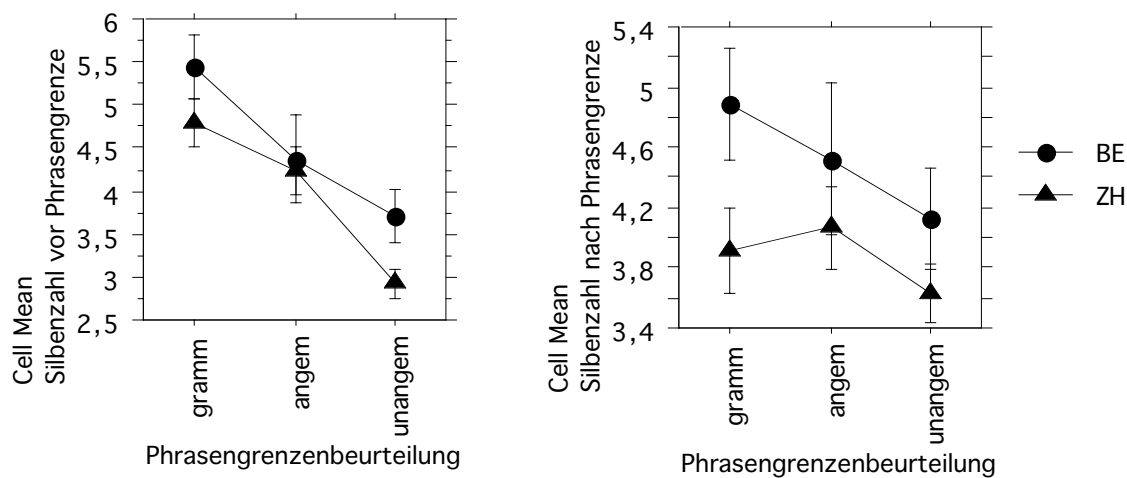


Abbildung 2: Mittelwert und 95%-Konfidenzintervall der Phrasenlänge in Silben vor der Phrasengrenze (links) und nach der Phrasengrenze (rechts) nach Beurteilung der Phrase und nach Dialekt

Die nachfolgende Phrasengrenze beeinflusst die Phrasendauer mehr als die vorangehende. Der auffälligste Unterschied zwischen den Sprechern liegt in der Phrasendauer nach grammatisch motivierter Grenze, die sich beim Zürcher nicht von den andern unterscheidet, während beim Berner nach grammatisch motivierten Grenzen längere Phrasen folgen. Abbildung 3 zeigt die Verteilung der Phrasenlängen in Abhängigkeit von der Grenzumgebung. Bei beiden Sprechern sind Phrasen, die auf ein grammatisches Wort enden, signifikant kürzer als solche, die mit einem lexikalischen Wort enden. Beim Zürcher Sprecher sind diese Phrasen signifikant kürzer als beim Berner Sprecher. Da die grammatischen Wörter im Allgemeinen kurz sind – Artikel, Pronomen, Konjunktionen und Präpositionen sind selten länger als zwei Silben –, deutet das auf eine unterschiedliche Phrasierungskonzeption hin: Der Zürcher Sprecher markiert relativ häufig eine Phrasengrenze nach einem einzelnen grammatischen Wort, wo nach dem Modell für die Lesesprache keine Pause stehen durfte. Grenzen bieten in psycholinguistischer Sicht einen Hinweis auf eine kognitive Belastung und stellen damit einen spontanen Ausdruck von Planungsprozessen dar oder aber sie haben kommunikative Bedeutung, indem sie ikonisch kognitive Entscheidungsprozesse signalisieren, wie das K. Fischer (1999) postuliert. Beim Zürcher Sprecher deutet die Kürze der vorangehenden Phrase bei diesen Grenzen auf diese zweite Kategorie hin. Die mit

grammatischen Wörtern endenden Phrasen des Berner Sprechers sind länger, was die Interpretation als tatsächliche Planungspausen etwas wahrscheinlicher macht.

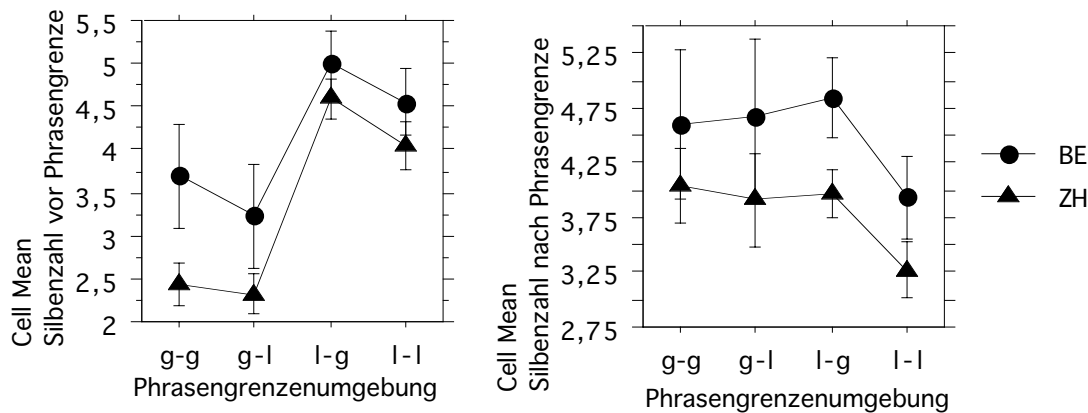


Abbildung 3: Mittelwert und 95%-Konfidenzintervall der Phrasenlänge vor der Phrasengrenze (links) und nach der Phrasengrenze (rechts) nach Phrasengrenzenumgebung (g = grammatisches Wort, l = lexikalisches Wort) und Dialekt

Nach Grenzen sind alle Phrasen etwa gleich lang, außer diejenigen zwischen zwei lexikalischen Wörtern, was noch geklärt werden muss. Hier besteht kein Unterschied zwischen den Sprechern.

## 6.2 Berner und Zürcher Timing

Unter Timing wird die zeitliche Strukturierung auf Segmentebene verstanden. Bevor einzelne Lautklassen angesehen werden, soll ein Blick auf die Silbenstruktur und das Vorkommen der Laute geworfen werden. So kann ausgeschlossen werden, dass Unterschiede im Timing auf eine unterschiedliche Lautverteilung oder Silbenstruktur zurückzuführen sind.

Die Berner Aufnahme zeigt mit 12.12% VC-Silben einen leicht höheren Anteil an Silben ohne Onset als die Zürcher Aufnahme mit 9.61%. Der Zusammenhang ist zwar signifikant aber sehr schwach<sup>9</sup>. Da das Schweizerdeutsche keinen Knacklaut zeigt, der für die Silbifizierung der deutschen Standardsprache wesentlich ist (D. Nubling und R. Schrambke in diesem Band), sind VC-Silben überhaupt möglich. Sie sind aber relativ selten, weil eine generelle Tendenz zur CV-Silbe besteht. Diese Tendenz ist im Schweizerdeutschen anders ausgeprägt als in der Standardsprache. Während in dieser silbenfinale Konsonanten von kopflosen Folgesilben im nächsten Wort oder Morphem durch einen den Onset bildenden Knacklaut geschieden werden, sind die Wort- und Morphemgrenzen in den schweizerdeutschen Mundarten weniger markiert. Der trennende Knacklaut fehlt, und so werden silbenfinale Konsonanten der Vorsilbe auch über die Wort- und Morphemgrenze hinaus als Onset an die ursprünglich vokalisches beginnende Silbe gebunden.

Konsonanten und Vokale sind in beiden Aufnahmen nur annähernd gleich häufig. In der Berner Aufnahme machen Vokale 34.66% aller Segmente aus, in der Zürcher Aufnahme sind es 32.52%. Die Korrelation ist signifikant, aber sehr schwach.<sup>10</sup>

<sup>9</sup> Kontingenzquotient = 0.041, Chi-Quadrat-p = 0.0004

<sup>10</sup> Kontingenzquotient = 0.02, Chi-Quadrat-p = 0.0012.



Die Vokalklassen sind in beiden Mundarten etwa gleich verteilt. Der auffälligste Unterschied zeigt sich bei Schwa. Dieses ist in Zürich rund 4% häufiger als in Bern, größere Unterschiede sind nur in den selteneren Klassen bei Diphthongen, Falldiphthongen, bei  $\emptyset / \text{æ}$ , den silbischen Konsonanten und Nasalvokalen zu sehen. Bei den Konsonanten sind die Unterschiede zwischen den beiden Aufnahmen noch geringer. Einzig die Halbvokale zeigen einen bedeutenden Unterschied, der auf die Vokalisierung von /l/ im Berndeutschen zurückzuführen ist. Somit wird deutlich, dass Unterschiede in der Prosodie nicht auf ein grundsätzlich anderes Verhältnis der einzelnen Laute zurückzuführen sind, sondern dass diese Differenzen in anderen, spezifisch prosodischen, Bereichen zu suchen sind, wie der zeitlichen und Intonationssteuerung.

Um die Unterschiede in der Segmentdauer zwischen den Dialekten darzustellen, werden Varianzanalysen durchgeführt. Diese verlangen nach normalverteilten Daten. Das folgende Histogramm mit der Dauer der einzelnen Segmente in ms zeigt aber eine stark linkssteile Verteilung (Abbildung 4 links). Es hat sich gezeigt, dass bei Segmentdauern von Sprachdaten die Berechnung im logarithmischen Raum zu einer Normalverteilung führt, das ist auch hier der Fall, wie aus (Abbildung 4 rechts) ersichtlich ist.

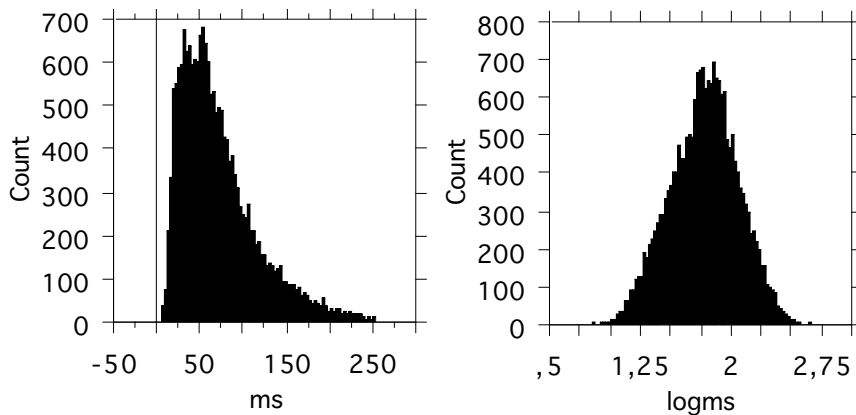


Abbildung 4: Histogramm der Segmentdauern in ms (links) in log ms rechts

Die Arbeit an der französischen und an der deutschen Sprachsynthese am Laboratoire d'Analyse Informatique de la Parole an der Universität Lausanne hat gezeigt, dass die intrinsische Segmentdauer einen großen Teil der Gesamtvariation erklären kann. Als erstes soll deshalb die Dauer einzelner ausgewählter Phonemklassen dargestellt werden.

Es wurde schon darauf hingewiesen, dass der Zürcher Sprecher bedächtiger spricht als der Berner Sprecher. Der Horeindruck bestätigt sich auch für die Segmentdauern, die beim Zürcher Sprecher entgegen dem gängigen Stereotyp im Durchschnitt länger sind als beim Berner Sprecher. Abbildung 5 zeigt, dass die Längenunterschiede vor allem bei Vokalen vorkommen, während die Konsonanten zwischen den Sprechern ähnlicher sind.

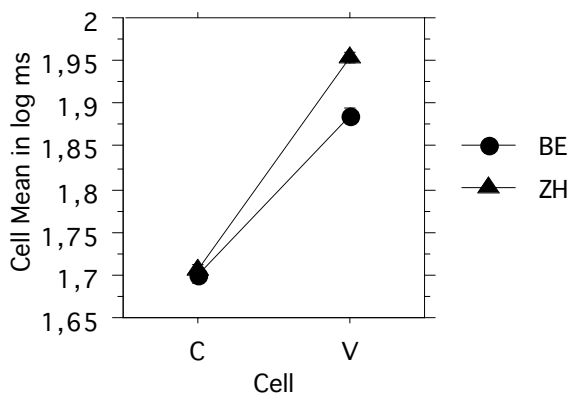


Abbildung 5: Mittelwert und 95%-Konfidenzintervall der Segmentdauer in log ms nach Lautklasse (Vokal / Konsonant) und nach Dialekt<sup>11</sup>

Abbildung 6 zeigt die Unterschiede in der Realisierung von Konsonanten und Vokalen nach Akzentuierung. Unterschieden werden Silben mit Schwa, unakzentuierte Silben mit Vollvokal und akzentuierte Silben mit Vollvokal. Wieder sind die Unterschiede zwischen den beiden Sprechern im Bereich der Konsonanten nicht sehr bedeutsam. Beide realisieren Konsonanten in Schwa-Silben hoch signifikant kürzer als in Silben mit Vollvokal. Konsonanten in akzentuierten und nicht akzentuierten Silben werden nicht unterschieden. Anders sieht es bei den Vokalen aus: Während Schwach bei beiden Sprechern gleich lang sind, sind die Längenunterschiede zwischen akzentuierten und nicht akzentuierten Silben jeweils signifikant. In beiden Fällen sind die Realisierungen durch den Zürcher Sprecher auch signifikant länger als die des Berner Sprechers.

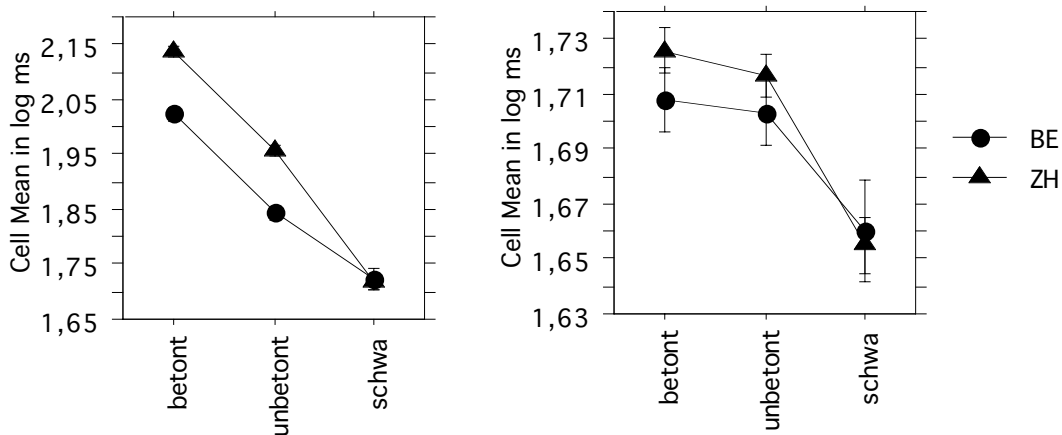


Abbildung 6: Mittelwert und 95%-Konfidenzintervall der Vokaldauer (links) und Konsonantendauer (rechts) nach Akzentuierung und nach Dialekt<sup>12</sup>

Die Unterscheidung in lexikalische und grammatische Wörter ist in den meisten Sprachen ein wichtiger Faktor für die Segmentdauer. Dies trifft auch für die beiden Mundarten zu: Wie

<sup>11</sup> Der Unterschied zwischen dem Zürcher und dem Berner Sprecher ist im Bereich der Konsonanten mit einem p-Wert von 0.052 knapp nicht signifikant, während er bei den Vokalen mit  $p < 0.001$  hochsignifikant ist. – Zur Verdeutlichung die Werte in ms. BE: C = 61 ms, V = 89 ms; ZH: C = 60 ms, V = 106 ms.

<sup>12</sup> t-tests für die Unterschiede zwischen Berner und Zürcher Sprecher in Schwa-Silben: Konsonantendauer  $p = .60$ ; Vokaldauer  $p = .54$ . t-test für die Unterschiede zwischen Berner und Zürcher Sprecher in unbetonten Silben: Konsonantendauer  $p = .0490$ ; Vokaldauer  $p < .0001$ . t-test für die Unterschiede zwischen Berner und Zürcher Sprecher in betonten Silben: Konsonantendauer  $p = .0090$ ; Vokaldauer  $p < .0001$ . – Zur Verdeutlichung die Werte in ms: Vokale: BE: betont = 116 ms, unbetont = 79, schwa = 60 ms; ZH: betont = 147 ms, unbetont = 103, schwa = 60 ms; Konsonanten: BE: betont = 63 ms, unbetont = 62, schwa = 56 ms; ZH: betont = 63 ms, unbetont = 62, schwa = 53 ms

üblich sind Segmente in Funktionswörtern kürzer als solche in lexikalischen Wörtern. Die Unterschiede sind beim Berner Sprecher ausgeprägter als beim Zürcher. Über die generelle Tendenz hinaus zeigen sich feinere Unterschiede, wenn die Akzentuierung mitberücksichtigt wird. Aus Abbildung 7 lässt sich herauslesen, dass das Verhältnis der Vokaldauern in lexikalischen und grammatischen Wörtern von beiden Sprechern je nach Betonung anders gestaltet wird.

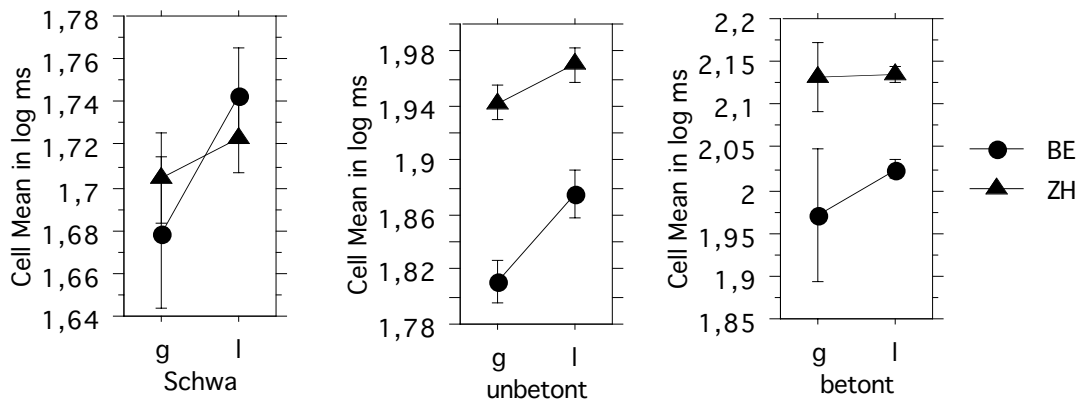


Abbildung 7: Mittelwert und 95%-Konfidenzintervall der Vokaldauer (Schwa links, unbetont Mitte, betont rechts) in log ms nach grammatischen / lexikalischen Wörtern und nach Dialekt<sup>13</sup>

Der Berner Sprecher unterscheidet die Segmentdauern zwischen den beiden Wortklassen stärker als der Zürcher Sprecher. Die Unterschiede zwischen den Wortklassen sind beim Berner Sprecher bei Schwa- und unbetonten Silben hoch signifikant, bei betonten Silben signifikant. Für den Zürcher Sprecher ist der Unterschied nur bei den unbetonten Silben signifikant. Dass die Werte beim Zürcher Sprecher zusammenfallen, kann mit der sorgfältigeren Artikulation zu tun haben. Dadurch sind Reduktionen weniger ausgeprägt, die sonst beim schnellen Sprechen üblicherweise vorkommen. Dieses Phänomen trifft, weil vor allem Funktionswörter reduziert werden, vor allem diese grammatischen Wörter.

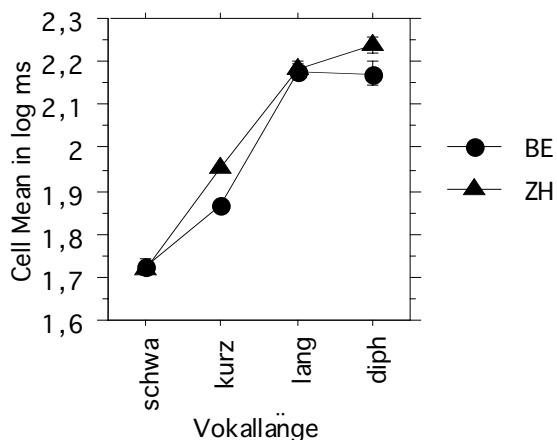


Abbildung 8: Mittelwert und 95%-Konfidenzintervall der Segmentdauer der Vokale in log ms nach phonologischer Länge und nach Dialekt<sup>14</sup>

<sup>13</sup> Zur Verdeutlichung Werte in ms: schwa: BE: g = 55 ms, l = 63 ms; ZH: g = 58 ms, l = 60 ms; unbetont: BE: g = 72 ms, l = 86 ms; ZH: g = 100 ms, l = 106 ms; betont: BE: g = 104 ms, l = 116 ms; ZH: g = 146 ms, l = 147 ms.

<sup>14</sup> Die Unterschiede zwischen den einzelnen Klassen sind jeweils hochsignifikant ( $p < 0.001$ ). Nur Langvokale und Diphthonge beim Berner Sprecher unterscheiden sich nicht signifikant. Zur Verdeutlichung Werte in ms: BE: diph = 161 ms, lang = 162 ms, kurz = 80 ms, schwa = 60 ms; ZH: diph = 187 ms, lang = 163 ms, kurz = 99 ms, schwa = 60 ms

Abbildung 8 zeigt die Verteilung der Vokaldauer nach phonologischer Längedistinktion. Der Zürcher Sprecher zeigt vier verschiedene Klassen, während der Berner Sprecher Diphthonge und Langvokale nicht unterscheidet. Beim Vergleich zwischen den Sprechern sind die Unterschiede bei den Kurzvokalen und bei den Diphthongen signifikant. Schwas und Langvokale unterscheiden sich nicht.

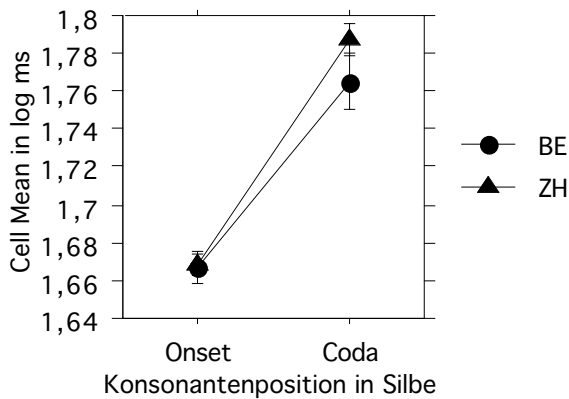


Abbildung 9: Mittelwert und 95%-Konfidenzintervall der Konsonantendauer in log ms nach Position in der Silbe und nach Dialekt<sup>15</sup>

Zum Schluss sollen noch Aspekte im Konsonantismus angesehen werden. Abbildung 9 zeigt die Länge der Konsonanten nach Position in der Silbe. Auch im Schweizerdeutschen sind die Konsonanten im Onset kürzer als diejenigen in der Silbencoda, wie im Französischen (E. Keller und B. Zellner 1996) oder in der deutschen Standardsprache (M. Riedi 1998, 52; B. Siebenhaar et al. 2001, 169). Der Unterschied zwischen Onset und Coda ist beim Zürcher Sprecher deutlicher als beim Berner Sprecher. Die beiden Sprecher unterscheiden sich nicht im Bezug auf die Dauer der Konsonanten im Onset, der Unterschied in der Coda ist jedoch hoch signifikant.

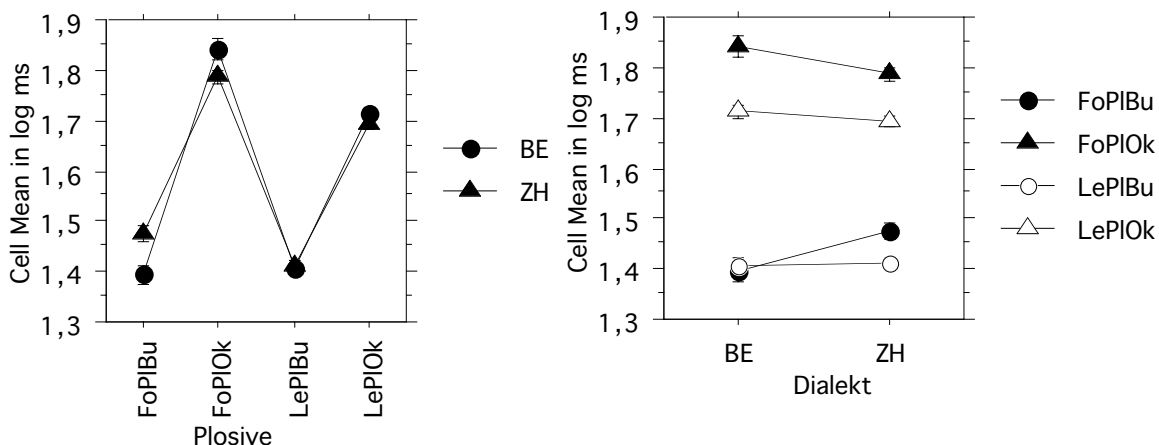


Abbildung 10: Mittelwert und 95%-Konfidenzintervall der Dauer der Plosive in log ms nach Fortes = Fo / Lenes = Le (je Ok = Okklusion und Bu = Burst) und nach Dialekt<sup>16</sup>

Abbildung 10 zeigt - in zwei unterschiedlichen Darstellungen - die Dauer der Plosive, und zwar Lenes und Fortes getrennt und Okklusion und Burst getrennt. Bei Lenis und Fortis ist

<sup>15</sup> Zur Verdeutlichung Werte in ms: BE: Onset = 55 ms, Coda = 75 ms; ZH: Onset = 54 ms, Coda = 73 ms.

<sup>16</sup> Zur Verdeutlichung Werte in ms: BE: FoPIBu = 28 ms, FoPIOk = 84 ms, LePIBu = 28 ms, LePIOk = 57 ms; ZH: FoPIBu = 35 ms, FoPIOk = 69 ms, LePIBu = 29 ms, LePIOk = 54 ms

die Okklusion länger als der Burst. Der Zürcher Sprecher unterscheidet sich vom Berner Sprecher in allen Bereichen außer beim Burst der Lenes (leerer Kreis). Die Okklusionsphase des Zürcher Sprechers ist bei Fortes und Lenes signifikant kürzer (Dreiecke), während der Burst bei Fortes länger ist als beim Berner Sprecher (schwarzes Dreieck)<sup>17</sup>. Interessant ist vor allem das Verhältnis von Bursts und Okklusion zwischen Fortes und Lenes. Während der Zürcher Fortes und Lenes nicht nur in der Okklusion unterscheidet, sondern auch im Burst, sind die Bursts beim Berner Sprecher nicht unterschieden, dagegen ist der Unterschied der Okklusions deutlicher. Somit unterscheiden beide Sprecher die Gesamtlänge der Plosive nicht, jedoch ist das Verhältnis von Okklusion und Burst unterschiedlich gestaltet.

## 7 Schluss

Die Datenanalyse ist noch nicht abgeschlossen, und hier konnte nur ein Ausschnitt gegeben werden. Aus dem bisher Dargestellten kann aber gezeigt werden, dass sich der Rhythmus des Berndeutschsprechers vom Rhythmus des Zurichdeutschsprechers unterscheidet. Das zeigt sich in der Phrasierung einerseits als unterschiedliche durchschnittliche Phrasenlänge, andererseits an der unterschiedlichen Verteilung Phrasenlänge je nach Art der Phrasengrenze. Im Timing sind die Unterschiede im Bereich der Vokale größer als bei Konsonanten. Aber auch für Konsonanten, hier am Beispiel der unterschiedlichen Verhältnisse bei Plosiven, konnten Differenzen gezeigt werden. Sehr häufig bewegen sich diese Unterschiede im Rahmen von einigen wenigen Millisekunden. Da aber schon zeitliche Veränderungen von 2% wahrgenommen werden können, wird die Summe dieser Unterschiede einen Teil des hörbaren Unterschieds von Berndeutsch und Zurichdeutsch ausmachen. Problematisch an den dargestellten Daten ist, dass aus der Analyse zweier Sprecher noch kaum Verallgemeinerungen möglich sind.

Der Einblick in diese Analyse hat auch gezeigt, dass für die Untersuchung der Prosodie nicht nur Intonationsmuster untersuchungswürdig sind, sondern dass auch die zeitliche Steuerung ein Feld ist, das es zu beackern lohnt.

Für die weitere Arbeit zu einer Mundartsynthese wird die Analyse weitergeführt und muss für jede Mundart in ein Modell der Phrasierung und der Segmentdauer führen. Zudem müssen Modelle für die Intonation erstellt werden. Nach der Umsetzung in Algorithmen, sowie der Fertigstellung der Diphondatenbank und der neuen Signalgeneration durch Eric Keller werden wir die Möglichkeit haben Mundart vom Computer vorlesen zu lassen, Modelle zu vergleichen und zu verändern.

## Literaturverzeichnis

- AUER, Peter / GILLES, Peter / PETERS, Jörg / SELTING, Margret (2000): Intonation regionaler Varietäten des Deutschen. In: STELLMACHER, Dieter (Hg.): Dialektologie zwischen Tradition und Neuansätzen. Stuttgart: Steiner (ZDL Beihefte 109), 222–239.
- CAMPBELL, W. N. (1997): Synthesizing Spontaneous Speech. In: SAGISAKA, Y. / CAMPBELL, N. / HIGUCHI, N. (eds.): Computing Prosody, New York: Springer, 165–186.

---

<sup>17</sup> Vgl. zu Fortis und Lenis im Zurichdeutschen Urs Willi (1996).

- FISCHER, Kerstin (1999): Die Ikonizität der Pause: Zwischen kognitiver Last und kommunikativer Funktion. In: WACHSMUTH, Ipke / JUNG, Bernhard (eds.), *KogWis99: Proceedings der 4. Fachtagung der Gesellschaft für Kognitionswissenschaft*, Bielefeld, 28. September - 1. Oktober 1999. Sankt Augustin: Infix, 250–255.
- FITZPATRICK-COLE, Jennifer (1999): The alpine intonation of Bern Swiss German, In: OHALA, John (ed.): *Proceedings of the XIVth International Congress of Phonetic Sciences (ICPhS)*, San Francisco. 941–944.
- FUJISAKI, Hiroya / HIROSE, Keikichi (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.
- GEE, J. and GROSJEAN, F[ranois] (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology* 15, 411–458.
- GIBBON, Dafydd (1998): Intonation in German. In: HIRST, Daniel / DI CRISTO, Albert: *Intonation Systems*. Cambridge: Cambridge University Press. 78–95.
- GILLES, Peter (2001): Die Intonation der Weiterweisung im Hamburgischen und Berlinischen. In: *Zeitschrift für germanistische Linguistik* 29, 40–69.
- GILLES, Peter / SCHRAMBKE, Renate (2000): Divergenz in den Intonationssystemen rechts und links des Rheins. In: FUNK, Edith / KÖNIG, Werner / RENN, Manfred (Hg.): *Bausteine zur Sprachgeschichte*. Heidelberg: Winter, 87–98.
- GROSJEAN, F[ranois] / COLLINS, M. (1979). Breathing, pausing, and reading. In: *Phonetica*, 36, 98–114.
- HIRST, Daniel and DI CRISTO, Albert (eds.) (1998): *Intonation Systems*. Cambridge: Cambridge University Press.
- KELLER, Eric (1994). Fundamentals of phonetic science. In: KELLER, Eric (ed.), *Fundamentals of Speech Synthesis and Speech Recognition*. Chichester: John Wiley. 5–21.
- KELLER, E[ric] / ZELLNER, B[rigitte] / WERNER, S[tefan] / BLANCHOU, N. (1993). The prediction of prosodic timing: Rules for final syllable lengthening in French. In: *Proceedings ESCA Workshop on Prosody September 27–29*. Lund, Sweden. 212–215.
- KELLER, Eric / ZELLNER, Brigitte (1995): A statistical timing model for French. In: *Proceedings of the XIIIth International Congress of the Phonetic Sciences*. Stockholm. 302–305
- KELLER, Eric / ZELLNER, Brigitte (1996). A timing model for fast French. In: *York Papers in Linguistics (University of York)* 17, 53–75.
- KOHLER, K[laus] J. (1997): Modelling Prosody in Spontaneous Speech. In: SAGISAKA, Y. / CAMPBELL, N. / HIGUCHI, N. (eds.): *Computing Prosody*, New York: Springer, 187–210.
- MIXDORFF, Hansjörg (2002): *An Integrated Approach to Modeling German Prosody*. Dresden: w.e.b Universitätsverlag (Studententexte zur Sprachkommunikation 25).
- MONNIN, P. / GROSJEAN, F[ranois] (1993). Les structures de performance en français: caractérisation et prediction. In: *L'Année Psychologique* 93, 9–30.
- RIEDI, Marcel P. (1998): *Controlling Segmental Duration in Speech Synthesis Systems*. Zurich (TIK-Schriftenreihe 26).
- SELTING, Margret (1995): *Prosodie im Gespräch*. Tübingen: Niemeyer (Linguistische Arbeiten 329).
- SIEBENHAAR, Beat (i. Dr.) *Sprachsynthese als Methode für die Dialektologie*. In: SCHEURINGER, Hermann / GAISBAUER Stephan: *Tagungsberichte der 8. Bayerisch-österreichischen Dialektologentagung in Linz*, 19. – 23. September 2001. Linz (Schriften zur Literatur und Sprache in Oberösterreich).

SIEBENHAAR, Beat / FORST, Martin / KELLER, Eric (i.pr.): Speech synthesis of dialectal variants as a method for research on prosody. In: Tagungsband zur Eleventh International Conference on Methods in Dialectology.

SIEBENHAAR, Beat / ZELLNER KELLER, Brigitte / KELLER, Eric (2001): Phonetic and Timing Considerations in a Swiss High German TTS System. In: KELLER, Eric / BAILLY, Gerard / MONAGHAN, Alex / TERKEN, Jacques / HUCKVALE, Mark (eds.): Improvements in Speech Synthesis. Chichester: John Wiley. 165–175.

WILLI, Urs (1996): Die segmentale Dauer als phonetischer Parameter von "fortis" und "lenis" bei Plosiven im Zurichdeutschen. Stuttgart: Steiner (ZDL Beihefte 92).

ZELLNER, Brigitte (1994): Pauses and the temporal structure of speech. In: KELLER, Eric (ed.): Fundamentals of speech synthesis and speech recognition. Chichester: John Wiley. 41–62.

ZELLNER, Brigitte (1996): Structures temporelles et structures prosodiques en français lu. In: Revue Française de Linguistique Appliquée: La communication parlée 1. Paris. 7–23.

ZELLNER, Brigitte (1998): Caractérisation et prédiction du débit de parole en français. Une étude de cas. Thèse de Doctorat. Faculté des Lettres, Université de Lausanne.

ZELLNER KELLER, Brigitte (2002a). La simulation du rythme de parole. In: Travaux de l'Institut de Phonétique de Strasbourg 31, 139–165.

ZELLNER KELLER, Brigitte (2002b). Revisiting the Status of Speech Rhythm. In: BEL, Bernard / MARLIEN, Isabelle (eds.), 2002. Proceedings of the Speech Prosody 2002 conference, 11–13 April 2002. Aix-en-Provence: Laboratoire Parole et Langage. 727–730.

ZELLNER KELLER, Brigitte (i. pr.). Prediction of Temporal Structures for Various Speech Rates. In: Campbell, N. (ed): Volume on Speech Synthesis. New York: Springer.

ZELLNER KELLER, Brigitte, KELLER, Eric (2001). Representing Speech Rhythm. In: KELLER, Eric / BAILLY, Gerard / MONAGHAN, Alex / TERKEN, Jacques / HUCKVALE, Mark (eds.): Improvements in Speech Synthesis. Chichester: John Wiley & Sons. 154–164.

ZIMMERMANN, Gerhard (1998): Die 'singende' Sprechmelodie im Deutschen. In: Zeitschrift für germanistische Linguistik 26, 1–16.