Beat Siebenhaar (Bern und Lausanne)

Die Modellierung zeitlicher Strukturen im Schweizerdeutschen

1 Einleitung

Die Prosodie der Mundarten wurde schon früh als auffälliges und distinktes Merkmal wahrgenommen und in mehreren Arbeiten zur Grammatik des Schweizerdeutschen mittels Musiknoten festgehalten (u. a. J. Vetsch 1910, E. Wipf 1910, K. Schmid 1915, W. Clauss 1927, A. Weber 1948), wobei schon A. Weber (1948, S. 53) anmerkt, "dass sich der musikalische Gang der Rede nicht ohne Gewaltsamkeit mit der üblichen Notenschrift darstellen lässt". Da also eine adäquate Kodierung, eine theoretische Grundlage und die notwendigen phonetischen Instrumente zur Intonationsforschung fehlten, wurden diese ersten Ansätze nicht aus- und weitergeführt. Erst in der Mitte des 20. Jahrhunderts brachte die technische Entwicklung Instrumente zur Messung der Prosodie hervor, die nun durch die Popularisierung der entsprechenden Computerprogramme im Übergang zum 21. Jahrhundert für die linguistische Forschung intensiv und breit genutzt werden können.
Die ersten grundlegenden Arbeiten zur deutschen Prosodie von O. von Essen (1964) sowie von Isatschenko und Schädlich (1964) zur Perzeption von 'tone switches' und der Text in der Tradition der generativen Grammatik von M. Bierwisch (1966) orientierten sich vor allem an der Standardsprache. Seit der Mitte der 1990er Jahre beschäftigt sich die Forschung nun aber vermehrt auch mit der regionalen Intonation. Ein wesentlicher Anstoß ist von M. Seltings (1995) kommunikationsanalytischer Arbeit ausgegangen, die nicht mehr auf isolierten Sätzen, sondern auf Gesprächssequenzen basiert. Dadurch wurde die traditionelle Bindung an die Einsatz-Syntax gelöst und die Systematik spontaner Daten wichtiger. Der von M. Selting und zuvor schon von C. Féry (1993) und S. Uhmann (1991) für das Deutsche adaptierte Ansatz der autosegmentalen Prosodie, wie sie für das Englische von J. Pierrehumbert (1980) entwickelt worden ist, ist in der Folge für die Beschreibung spontansprachlicher, regionaler und dialektaler Daten verwendet worden. Seit 2000 ist in diesem Forschungsparadigma eine größere Anzahl Publikationen zur regionalen Intonation des Deutschen erschienen (vgl. insbesondere die Arbeiten von P. Auer, P. Gilles, J. Peters und M. Selting).
Ein zweiter Strang der Prosodieforschung geht von der technologischen Forschung zur Sprachsynthese und Spracherkennung aus, wo die Prosodie als wesentlich für die semantische und grammatische Strukturierung von Äußerungen erkannt wurde. Zudem ist weitgehend anerkannt, dass die Prosodie für die Natürlichkeit und damit die Akzeptabilität der Sprachsynthese einen wesentlichen Anteil ausmacht. Im Abschlussband (E. Keller et al. 2001) zum COST 258-Projekt, das die Natürlichkeit der Sprachsynthese zum Thema hatte, machen die Teile zur Prosodie rund einen Drittel aller Beiträge aus. In wesentlichen Teilen ist diese technologische Forschung darauf ausgerichtet eine möglichst hohe Korrelation eines Modells mit realen Daten zu erreichen. Von linguistischer Seite werden diese Daten kaum genutzt. Einerseits werden aus den für die technologischen Bedürfnisse erstellten parametrischen Regressionsmodellen meist keine linguistischen Schlussfolgerungen gezogen, die durchaus möglich wären[1]. Die immer häufiger verwendeten nichtparametrischen Regressionen oder künstlichen neuronalen Netzwerke (alltagsweltlich häufig als künstliche Intelligenz bezeichnet) auf der anderen Seite bieten kaum Möglichkeiten der linguistischen Analyse. Da überdies ein großer Teil der Forschung privat ist, sind die Resultate für die Linguistik nicht zugänglich. Die kommerzielle Ausrichtung dieser Forschung legt zudem den Schwerpunkt

[1] Siehe dazu als Beispiel für viele andere B. Möbius / J. van Santen (1996) oder im Vergleich der Modelle verschiedener Sprachen bei J. van Santen (1998).

auf die Standardsprachen, da eine dialektale und damit regionale Ausrichtung die Marktfähigkeit eines Produkts einschränkte. Eine Ausrichtung auf die regionale Variante des Deutschen in der Schweiz findet sich in B. Siebenhaar et al. (2001).
Die Forschung und die Konzepte der Sprachsynthese sollen im Projekt, in das die hier vorgestellte Arbeit eingebettet ist[2], für die linguistische Erforschung der mundartlichen Prosodie nutzbar gemacht werden. Im Gegensatz zu rein analytischen Zugängen, deren Schwerpunkt auf der Beschreibung und Erklärung einzelner spezifischer Intonationskonturen oder typischer intonatorischer Realisierungen kommunikativer Funktionen liegt, geht die auf eine Synthese ausgerichtete Methode zur Prosodieforschung in einer holistischen Konzeption von globalen und generellen Aspekten aus, die im Lauf der Analyse und Modellierung spezifiziert werden. Damit zwingt die Synthese dazu, alle Aspekte der Sprache zu beachten, auch diejenigen, die mit analytischem Vorgehen beiseite gelassen werden können. Eine solche Ausrichtung auf die Synthese besagt aber nicht, dass auf die analytischen Zugänge verzichtet wird oder dass sie vernachlässigt werden. Es ist vielmehr so, dass die Analyse nicht Selbstzweck ist, sondern dass sie auf die Synthese ausgerichtet ist. In diesem Sinne sind auch die folgend dargestellten Modellierungen der Timingstrukturen zweier Mundarten zu verstehen.

2 Timing innerhalb der Prosodieforschung

Die neuere Prosodieforschung konzentriert sich auf die Analyse der Intonation, während die Phrasierung, das Timing und auch die Intensität viel weniger Beachtung finden. Die Stimmqualität als prosodisches Element, z. B. Knarrstimme zur Markierung von Phrasenenden, ist bisher überhaupt noch nicht untersucht worden. Diese Gewichtung innerhalb der Forschung trifft sowohl auf die Arbeit mit der Standardsprache als auch auf die variationslinguistischen bzw. dialektologischen Arbeiten zu.
Die Phrasierung, die Strukturierung von Sätzen und Äußerungen in Teile mit höherer Kohäsion und Teile mit niedriger Kohäsion wird vor allem im Zusammenhang mit der Syntax und im Hinblick auf psycholinguistische und kognitive Aspekte untersucht. Forschung zum Timing, d. h. zur temporalen Struktur von Sprache, steht häufig in einem phonologischen Kontext. Aktuell sind Fragen zu phonetischen Korrelaten von phonologischen Distinktionen, wie Vokalquantität (S. Schmid, i. Dr.), Fortis-Lenis (U. Willi 1996), Geminaten (W. Ham 2001), Silbenschnitt (H. Spiekermann 2000, P. Auer et al. 2002) und Rhythmisierung (Neuber 1998/i. Dr.). Hinzu kommen Modellierungen im Zusammenhang mit der Sprachsynthese, welche die Dauer der einzelnen Segmente[3] vorhersagen. Dabei stehen sich Modelle gegenüber, welche die Silbenlänge erklären und die Segmentlänge davon ableiten, und solche, die die Segmentlänge direkt erarbeiten. Als dritte Kategorie versucht ein nicht-segmentaler Ansatz (R. Ogden et al. 1999), die zeitliche Steuerung in den akustischen Merkmalen im Sprachsignal – Stimmhaftigkeit, Nasalität, Formantübergänge – zu modellieren, die nicht direkt an das Segment gebunden sind. Jedes dieser Modelle zeigt ein unterschiedliches Set von Inputfaktoren. Die meisten Modelle berücksichtigen Faktoren auf Segment-, auf Silben-

[2] Das Projekt "Erarbeitung von Grundlagen zur Erforschung schweizerdeutscher Prosodie mittels sprachsynthetischer Modellierung" ist vom Schweizerischen Nationalfonds finanziert.
[3] In der Sprachsynthese hat sich der Terminus Segment etabliert, der anstelle von Phonem verwendet wird. Dabei greift der Begriff teilweise weiter, teilweise weniger weit als Phonem. Beispielsweise werden – wie in den vorliegenden Modellen – häufig Okklusionsphase und Burst von Plosiven als zwei Segmente betrachtet, oder die silbischen Konsonanten, die im Deutschen phonologisch Schwa + Konsonant entsprechen, werden als eigene Segmente betrachtet.

und auf Phrasenebene. Auf Segmentebene sind das meist eine intrinsische Länge und Einflüsse der benachbarten Segmente, sowie die Position in der Silbe. Auf Silbenebene werden Akzentuierung und die Silbenposition in Akzentgruppen berücksichtigt. Auf Phrasenebene zeigt sich, dass die Position des Segments in Relation zu Phrasen- und Satzgrenzen einen Einfluss auf das Timing hat. Unterschiedlich sind auch die Ansätze, wie ein konkreter Dauerwert in ms aus den abstrakten Inputfaktoren errechnet wird. Neben regelgeleiteten Ansätzen sind vor allem statistische Ansätze verbreitet, wobei sowohl parametrische als auch nicht-parametrische Modelle vertreten werden. Eine Darstellung der verschiedenen für das Deutsche angewendeten Modelle findet sich in H. Mixdorff (2002, S. 27–37).

Die Erforschung des Timings war auch in der auf die Synthese ausgerichteten Forschung bisher im Allgemeinen weniger prominent als die Intonationsforschung. Da jedoch die Silbendauer in gelesenen Texten zwischen verschiedenen Sprechern viel stärker korreliert als der Grundfrequenzverlauf (E. Keller 1994, S. 7), wurde für die französische Synthese LAIPTTS_F eine primäre Ausrichtung am Timing vertreten (E. Keller / B. Zellner 1995, im Überblick in B. Zellner Keller 2002). Für die deutsche Synthese LAIPTTS_D ist dieses Konzept erfolgreich adaptiert worden (B. Siebenhaar et al. 2001). Dieser Ansatz stellt jedoch in der Syntheseforschung eine Ausnahme dar. Neueste Perzeptionstests zu deutschen Synthesen (C. Brinckmann / J. Trouvain 2003, H. Mixdorff / O. Jokisch 2003) haben jedoch die Bedeutung der zeitlichen Aspekte für die Qualität von Synthesen hervorgehoben. So halten H. Mixdorff und O. Jokisch (2003, S. 45) fest, dass "the accuracy of the predicted syllable durations appears to be a stronger factor with respect to the perceived quality than the accuracy of the predicted F0 contour". Auch wenn hier nicht direkt die Segment-, sondern die Silbendauern untersucht wurden, so wird trotzdem deutlich, dass der Zeitfaktor wesentlich ist. Da zudem eine adäquate Silbendauer von einer adäquaten Segmentdauer abgeleitet werden kann, trifft diese Aussage auch auf die Segmentdauern zu.

3 Eine Sprachsynthese für schweizerdeutsche Mundarten

Es wurde oben darauf hingewiesen, dass eine linguistische Methode, die auf eine Synthese ausgerichtet ist, andere Ergebnisse hervorbringen kann als rein analytische Zugänge (vgl. B. Siebenhaar i. Dr. a / B. Siebenhaar et al. i. Dr. a). So wird nun eine Synthese für zwei Deutschschweizer Mundarten – das Zürichdeutsche und das Berndeutsche – erstellt. Ziel ist dabei einerseits das Resultat, eine Synthese zweier Mundarten, die eine sprecherunabhängige Ausgabe verschiedener Dialekte ermöglicht, um diese für Perzeptionstests zur Verfügung zu haben. Die Sprachsynthese ist damit ein 'test-tool' für linguistische Fragen zur dialektalen Prosodie. Auf der anderen Seite ermöglicht die Analyse auf dem Weg zur Synthese Aussagen über die dialektale Prosodie. Somit ist es jetzt schon möglich, einzelne Resultate dieser Analyse zu präsentieren.

Sprachsynthesen lesen normalerweise einen geschriebenen Text vor. Eine Synthese mundartlicher Varianten kann dagegen nicht eine gelesene Sprache wiedergeben, weil die Dialekte kaum vorgelesen, sondern normalerweise spontan gesprochen werden. Deshalb muss einerseits der Input in die Synthese und andererseits der zu reproduzierende und damit der zu analysierende Sprechstil neu beurteilt werden. Da für die Schweizer Mundarten nicht auf eine normierte Orthographie zurückgegriffen werden kann, wird auf eine aufwendige und in jedem Fall inadäquate Graphem-Phonem-Übersetzung verzichtet und mit einer phonetischen Transkription als Input gearbeitet. Dies hat neben dem geringeren Programmieraufwand den Vorteil, dass das System flexibler ist, natürlich auf Kosten eines weniger attraktiven Inputs. Für die Analyse und somit für die Basis unserer Prosodiemodelle haben wir uns entschieden, den Stil des öffentlichen Interviews zu verwenden, das in den Schweizer Medien häufig oder

vermutlich sogar mehrheitlich in Mundart geführt wird. Interviews sind somit in eine natürliche Kommunikationssituation eingebettet, wir haben es also mit natürlichen Daten zu tun. Gleichzeitig ist diese Situation aber auch durch einen erhöhten Formalitätsgrad gekennzeichnet, was allzu 'exotische' prosodische Muster eher ausschließt.

4 Die Grundlagen für die Timing-Modelle

Für die Modellierung des Timings wurde für das Berndeutsche und das Zürichdeutsche je ein Sprecher aufgenommen[4]. Es wurden 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. Für den Berner Sprecher wurde ein Interview von gut 20 Minuten analysiert, für den Zürcher Sprecher sind es zwei Interviews von zusammen knapp 50 Minuten. Die Aufnahmen wurden mit Hilfe eines Aligners vorsegmentiert und manuell korrigiert[5]. Daraus ergaben sich für den Zürcher Sprecher 16'436 Segmente, für den Berner Sprecher sind es 8'462. Zusätzlich wurden Silben-, Wort-, Phrasen- und Satzgrenzen und der grammatische Status der Wörter als grammatische Wörter (= Artikel, Pronomen, Konjunktionen, Präpositionen), als Hilfsverben oder als lexikalische Wörter markiert. Diese Daten bilden nun die Grundlage für die Prosodiemodelle.
Stilistisch unterscheiden sich die beiden Sprecher insofern, als der Zürcher bedächtiger formuliert, jedoch weniger Selbstkorrekturen vornimmt, während der Berner weniger sorgfältig spricht und häufiger Selbstkorrekturen aufweist. Das äußert sich in den Daten so, dass der Berner Sprecher im Durchschnitt in einer Phrase 4.5 Silben aufweist, beim Zürcher Sprecher liegt dieser Durchschnitt bei nur 3.8 Silben pro Phrase. In B. Siebenhaar (i. Dr. b) wurde ein erster Versuch unternommen, diese unterschiedliche Phrasierung zu analysieren. Die weiterführende Arbeit hat aber gezeigt, dass die Analyse noch stark differenziert werden muss, um zu einer Modellierung der Phrasierung zu gelangen. Für die Analyse und Modellierung des Timings ist jedoch die Markierung der Phrasengrenzen im Transkript ausreichend, da deren Setzung in den gemessenen Daten gegeben ist und innerhalb des Timing-Modells nicht errechnet werden muss.
Für das Timing wurden relevante Faktoren mittels Varianzanalysen herausgearbeitet, die aus einer phonetischen Transkription herauszulesen sind – die Bedingung für eine Synthese – und welche die zeitliche Strukturierung bestimmen. Resultate dieser Analysen und Vergleiche der Sprecher finden sich in B. Siebenhaar (i. Dr. b) und in B. Siebenhaar et al. (i. Dr. b).
An dieser Stelle sollen nun nicht nochmals die Ergebnisse dieser Analysen verglichen werden, sondern es soll gezeigt werden, wie die in den Analysen als relevant erachteten Faktoren in die Timing-Modelle eingebettet werden. Damit wird auch deutlich, inwiefern sich die Modelle für das Berndeutsche und das Zürichdeutsche entsprechen und wo sich Unterschiede zeigen.
Wie unter Punkt 2 in der Darstellung der unterschiedlichen Modelle für die Sprachsynthese schon angesprochen, sind verschiedene Zugänge zur Timing-Modellierung aktuell. Das hier vorgestellte Modell errechnet die Segmentdauer direkt aus den gegebenen Inputfaktoren mittels eines parametrischen Regressionsmodells. Die dafür verwendete statistische Methode ist ein sogenanntes 'generalized linear model' (GLM). In diesem Verfahren wird die Segmentdauer als abhängige Variable betrachtet, welche mit den unabhängigen Variablen

[4] Für das Berndeutsche ist ein zweiter Sprecher analysiert worden, jedoch ist das Modell noch nicht ganz fertig.
[5] Vielen Dank an Martin Forst, Ingrid Hove und Katrin Häsler, die einen großen Teil dieser Arbeit geleistet haben.

mittels eines additiven Modells verbunden ist.[6] Die Dauer eines jeden Segments ist also bestimmt durch den Einfluss von Faktor 1 + Einfluss von Faktor 2 + … + Einfluss von Faktor n. In der Analyse wird das Gewicht der einzelnen Faktoren und Variablen berechnet; in der Synthese können für jedes Segment die Ausprägung dieser Faktoren bestimmt werden und dann die entsprechenden Werte eingesetzt und zusammengezählt werden.
Das analytische Vorgehen besteht in einer Varianzanalyse für einen als relevant erachteten Faktor. Wenn sich Variablen dieses Faktors signifikant unterscheiden, werden diejenigen Variablen mit ähnlichen Einflüssen zusammengefasst und nicht signifikante Variablen weggelassen, um das Modell stabiler zu gestalten. Anschließend bildet die bisher nicht erklärte Varianz den Input für die Varianzanalyse mit dem nächsten Faktor und das 'Spiel' beginnt von vorne. Das Verfahren wird unten exemplarisch durchgespielt. Die statistischen Analysen wurden mit GLMStat 5.7.4[7] durchgeführt.
Varianzanalysen verlangen nach normalverteilten Daten. Das Histogramm mit der Dauer der Segmente in ms zeigt eine stark linkssteile Verteilung (Abbildung 1 links). Für Segmentdauern von Sprachdaten hat sich gezeigt, dass die Berechnung im logarithmischen Raum annähernd zu einer Normalverteilung führt; das ist auch hier der Fall, wie aus (Abbildung 1 rechts) ersichtlich ist. Sämtliche Berechnungen werden deshalb mit log-transformierten Daten durchgeführt.

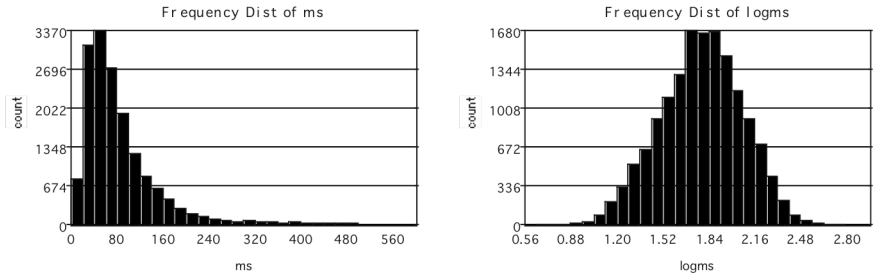

Abbildung 1: Frequenzverteilung der Segmentdauern in ms (links) und in log ms (rechts)

Die Grundlage für die hier entwickelten Modelle bilden die Faktoren, die für die französische Sprachsynthese als relevant erarbeitet wurden (E. Keller et al. 1993, E. Keller / B. Zellner 1996, B. Zellner 1998) und die für die Synthese des Schweizerhochdeutschen weitgehend adaptiert werden konnten (B. Siebenhaar et al. 2001). Mit diesen Modifikationen sind dies neben der intrinsischen Länge der einzelnen Segmente, die gemäß dem Ansatz von B. Zellner 1998 in Klassen ähnlicher Länge und ähnlicher Streuung zusammengefasst werden, der Einfluss der benachbarten Segmente, dann auf Silbenebene die Silbenstruktur und die Position des Segments in der Silbe. Auf Wortebene wird der grammatische Status des Worts, in dem das Segment vorkommt, berücksichtigt. Dann wird die Position der Silbe innerhalb des Wortes und innerhalb der Phrase berücksichtigt. Wie zu erwarten war, mussten die Variablen innerhalb dieser Faktoren neu gruppiert werden. Zudem wurde mit der Anzahl Konsonanten zwischen Vokalen ein zusätzlicher Faktor für die Vorhersage berücksichtigt.

5 Exemplarische Erstellung des Zürcher Timing-Modells

[6] Auf eine mathematische Herleitung wird hier verzichtet. Sie ist in den Handbüchern zu Statistiksoftware oder in spezifischen Einführungen zu GLM zu finden.
[7] GLMStat ist ein günstiges und einfach zu handzuhabendes Shareware-Programm für GLM von Ken Beath, das unter http://www.ozemail.com.au/~kjbeath/glmstat.html erhältlich ist. Die Anwendung von GLMStat für linguistische Analysen mit log-linearer Modellierung und logistischer Regression ist in J. Paolillo (i. Dr.) demonstriert.

Im Folgenden wird die Modellierung für das Timing der Zürcher Mundart dargestellt. Das Verfahren für das Berndeutsche entspricht diesem weitgehend.
Für die intrinsische Dauer der Segmente werden jeweils Segmente mit ähnlichem Dauer-Mittelwert und ähnlicher Streuung in Klassen zusammengefasst. Dabei ist zu bemerken, dass der Transkriptionsschlüssel für das Zürichdeutsche 115 Segmente umfasst. Der relativ große Umfang ist einerseits darauf zurückzuführend, dass die Schweizer Mundarten ein gegenüber der Standardsprache vor allem im Vokalismus viel größeres Phonemsystem haben, andererseits darauf, dass Plosive als zwei Segmente aufgefasst werden und dass silbische Konsonanten als eigenständige Segmente aufgenommen wurden. Vor allem von Bedeutung ist, dass lange und kurze Vokale je in akzentuierter und nicht akzentuierter Position unabhängig voneinander enthalten sind, weil die Länge und die Akzentuierung nicht alle Vokale linear gleich verändern. Diese 115 Segmente werden nun in Klasen mit ähnlichem Dauerverhalten zusammengefasst, dabei wird darauf geachtet, dass jede Klasse mindestens 150 Werte enthält, um die Stabilität des Modells zu gewährleisten. Für beide Mundarten wurden so je 12 Klassen herausgearbeitet; diese Klassen enthalten, weil für die Einteilung nur auf die statistische Streuung geachtet wird, jeweils ganz unterschiedliche Segmente. Klasse 5 für den Zürcher Sprecher enthält beispielsweise die Segmente {u, ʏ, l̩, n̩, lː, m, n, ŋː, Okklusion von b, t und p sowie Burst und Friktion von pf}.
Wenn nun diese 12 Klassen als unabhängige Variable für ein einfaches GLM verwendet werden, so erklärt dieses rund 56.64 % der gesamten Varianz. Abbildung 2 zeigt in der linken Grafik die Verteilung von gemessenen und mit diesem einfachen Modell wiedergegebenen Werten; die rechte Grafik zeigt den Zusammenhang von realen Segmentdauern und den Fehlern des Modells. Die Fehler haben definitionsgemäß einen Mittelwert von 0.0 log ms; die Standardabweichung beträgt 0.1861 log ms.

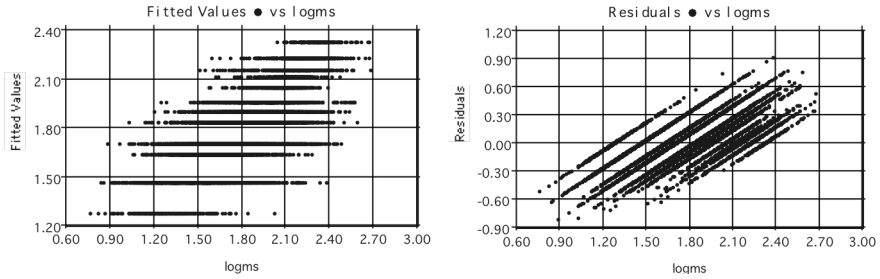

Abbildung 2: Scattergram: reale und modellierte Segmentdauern (links), reale Segmentdauern und Fehler (rechts). GLM: y = Lautklassen

In nächsten Schritt werden die Fehler des Modells als Input für ein erweitertes Modell genommen, und es wird versucht, diese Werte mit einem zusätzlichen Faktor zu erklären. Die Berücksichtigung des nachfolgenden Segments verbessert die erklärte Varianz um 5.81 %, die zusätzliche Berücksichtigung des vorangehenden Segments um 1.60 % und die Berücksichtigung des übernächsten Segments nochmals um 0.22 %. Für jeden dieser Schritte müssen zur Stabilisierung des Modells jeweils einzelne Variablen dieser Faktoren zusammengefasst werden. So finden sich für den Einfluss des vorangehenden Segments elf Variablen, für das folgende Segment neun Variablen, und für das übernächste Segment sind es noch drei. Mit diesen insgesamt vier Faktoren werden nun 64.27 % der Gesamtvarianz erklärt. Abbildung 3 zeigt – wie oben – in der linken Grafik die Verteilung von gemessenen und mit diesem 4-Faktoren-Modell errechneten Werten; die rechte Grafik zeigt den Zusammenhang von realen Segmentdauern und den Fehlern des Modells.

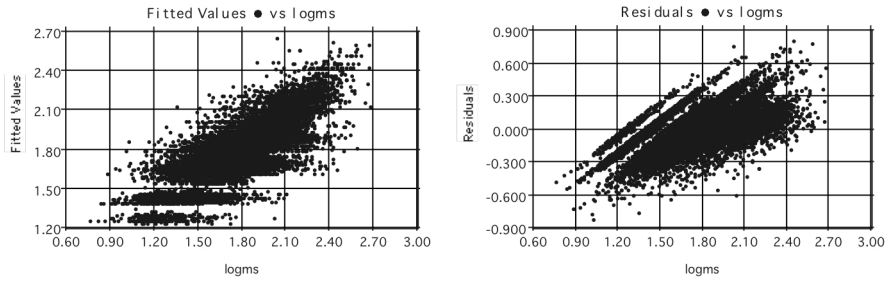

Abbildung 3: Scattergram: reale und modellierte Segmentdauern (links), reale Segmentdauern und Fehler (rechts). GLM: y = Lautklassen + vorangehendes Segment + folgendes Segment + übernächstes Segment

Als nächstes werden die Fehler des Modells mit silbenrelevanten Faktoren zu erklären versucht. In der Analyse haben sich die phonologische Länge des Nukleus (1.28 % Erklärung der Varianz), die Position des Segments in der Silbe (0.07 % Erklärung der Varianz[8]) und die Akzentuierung der Silbe, d. h. der Wortakzent, (0.22 % Erklärung der Varianz) als relevant erwiesen. Die Berücksichtigung der Anzahl Segmente im Onset bzw. in der Coda zeigte kaum eine signifikante Verbesserung des Modells, weshalb dieser Faktor weggelassen wurde, jedoch konnte mit der Anzahl Konsonanten zwischen Vokalen, ohne Berücksichtigung der Silbengrenze, eine Verbesserung um 0.26 % erzielt werden. Damit sind 66.1 % der Gesamtvarianz erklärt. Abbildung 4 dokumentiert die Verbesserung des Modells.

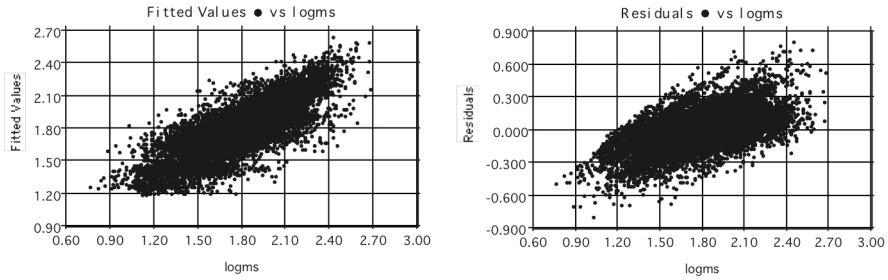

Abbildung 4: Scattergram: reale und modellierte Segmentdauern (links), reale Segmentdauern und Fehler (rechts). GLM: y = Lautklassen + vorangehendes Segment + folgendes Segment + übernächstes Segment + phonologische Länge des Nukleus + Position des Segments in der Silbe + Akzentuierung + Anzahl Konsonanten zwischen Vokalen

Wieder beginnt das Spiel von vorne mit Faktoren auf der Wortebene. Hier werden einerseits der grammatische Status des Wortes, was zusätzlich 0.15 % der Varianz erklärt, und andererseits die Position des Segments im Wort berücksichtigt, mit einer Verbesserung der Varianzerklärung um 0.02 %. Siehe dazu Abbildung 5.

[8] Der Anteil der Erklärung ist sehr gering, der Faktor wurde jedoch beibehalten, weil er im parallel entwickelten Modell für das Berndeutsche mit 0.68 % viel bedeutsamer ist.

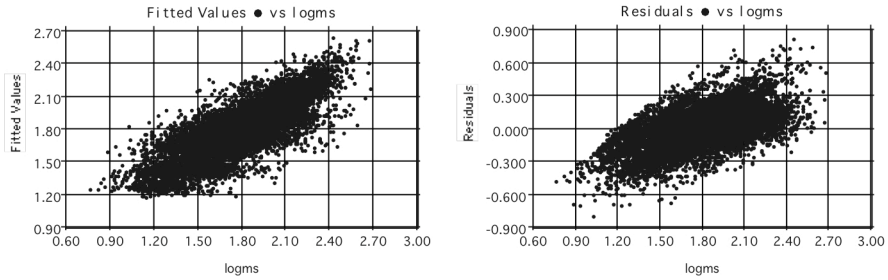

Abbildung 5: Scattergram: reale und modellierte Segmentdauern (links), reale Segmentdauern und Fehler (rechts). GLM: y = Lautklassen + vorangehendes Segment + folgendes Segment + übernächstes Segment + phonologische Länge des Nukleus + Position des Segments in der Silbe + Akzentuierung + Anzahl Konsonanten zwischen Vokalen + grammatischer Status des Wortes + Position des Segments im Wort

Zum Schluss werden noch Aspekte auf der Phrasenebene mitberücksichtigt. Der zusätzliche Faktor beachtet die Position der Silbe im Bezug auf eine große, bzw. eine kleine Phrasengrenze. Dieser Faktor verbessert das Modell nochmals um zusätzliche 1.1 %. Damit erklärt das Modell 67.4 % der Gesamtvarianz. Die Standardabweichung der Fehler beträgt 0.1615 log ms. Im Vergleich zum ersten Modell (s = 0.1861 log ms) zeigt die Fehlerstreuung also eine deutliche Reduktion.

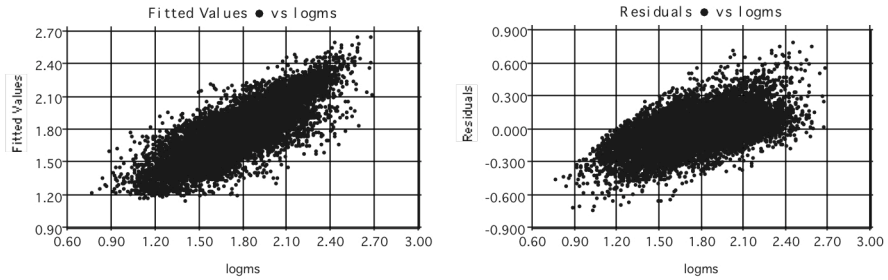

Abbildung 6: Scattergram: reale und modellierte Segmentdauern (links), reale Segmentdauern und Fehler (rechts). GLM: y = Lautklassen + vorangehendes Segment + folgendes Segment + übernächstes Segment + phonologische Länge des Nukleus + Position des Segments in der Silbe + Akzentuierung + Anzahl Konsonanten zwischen Vokalen + grammatischer Status des Wortes + Position des Segments im Wort + Position der Silbe in der Phrase

6 Der Vergleich der Modelle

Auf dieselbe Weise wurde ein Modell für das Berndeutsche erstellt. Es wurde darauf geachtet, dass für beide Modelle dieselben Faktoren zur Geltung kommen, um direkte Vergleiche zu ermöglichen. Tabelle 1 zeigt die Bedeutung der einzelnen Faktoren in den beiden Modellen.

Tabelle 1: Vergleich der Modelle für das Zürichdeutsche und das Berndeutsche. Die Prozentzahl der erklärten Varianz gibt den Wert für die additive Anwendung der Regeln (von oben nach unten in der Liste) an.

	Zürichdeutsch		Berndeutsch	
Effekt	erklärte Varianz	Verbes-serung	erklärte Varianz	Verbes-serung
Lautklassen	56.64 %		55.49 %	
folgendes Segment	62.45 %	5.81 %	59.41 %	3.92 %
vorangehendes Segment	64.05 %	1.60 %	60.52 %	1.11 %
übernächstes Segment	64.27 %	0.22 %	60.73 %	0.21 %
phonologische Länge des Nukleus	65.55 %	1.28 %	61.20 %	0.48 %
Position in der Silbe	65.62 %	0.07 %	61.89 %	0.68 %
Akzentuierung	65.84 %	0.22 %	61.94 %	0.05 %
Anzahl C zwischen V	66.10 %	0.26 %	62.24 %	0.30 %
Position des Segments im Wort	66.13 %	0.02 %	62.48 %	0.24 %
grammatischer Status des Wortes	66.28 %	0.15 %	62.59 %	0.11 %
Position der Silbe in der Phrase	67.42 %	1.14 %	63.45 %	0.86 %

Bei Interpretation und Vergleich der Modelle ist als erstes zu erwähnen, dass das Modell für das Berndeutsche die gemessenen Werte weniger gut wiedergeben kann als das Zürcher Modell. Das ist einerseits sicher darauf zurückzuführen, dass für das Zürichdeutsche ein fast doppelt so großes Corpus zur Verfügung steht wie für das Berndeutsche. Andererseits ist nicht auszuschließen, dass der Berner Sprecher eine größere Varianz aufweist als der Zürcher Sprecher, welche durch die zur Berechnung herangezogenen Faktoren nicht erklärt wird. Abgesehen davon macht der Blick auf die einzelnen Faktoren deutlich, dass diese nicht in beiden Modellen dieselbe Erklärungsrelevanz haben, das heißt also, dass wir es mit verschiedenen Modellen zu tun haben, was bedeutet dass sich die beiden Sprecher in der Strukturierung der zeitlichen Dimension von Sprache unterscheiden.
In beiden Modellen ist die intrinsische Länge des Segments bei Weitem der bedeutendste Faktor, der über die Hälfte der Gesamtvarianz erklärt. Dann folgen in beiden Modellen die umgebenden Segmente, wobei das folgende Segment ein um den Faktor 3.5 höheres Gewicht hat als das vorangehende Segment. Im Berner Modell haben diese beiden Faktoren nur je knapp 70 % der Erklärungsrelevanz des Zürcher Modells. Erst bei einer weiteren Differenzierung zeigen die Faktoren in den beiden Modellen eine unterschiedliche Gewichtung. Dabei fällt auf, dass die positionsbedingten Faktoren im Modell für das Berndeutsche jeweils wichtiger sind als im Zürcher Modell, wo deren Bedeutung mit Ausnahme der Phrasierung sehr gering ist. Im Zürcher Modell sind dagegen die Faktoren auf Silben- und Wortebene wichtiger. Das wird in der folgenden Abbildung 7 deutlich, wo die Pfeile angeben, wie sich Reihenfolge der Bedeutung der Faktoren innerhalb der Modelle ändert. Innerhalb dieser Gruppen ändert sich die Reihenfolge fast nicht, einzig der grammatische Status des Wortes und die Akzentuierung wechseln die Position.

Zürichdeutsch	
Effekt	Verbesserung
Lautklassen	56.64
folgendes Segment	5.81
vorangehendes Segment	1.6
phonologische Länge des Nukleus	1.28
Position der Silbe in der Phrase	1.14
Anzahl C zwischen V	0.26
übernächstes Segment	0.22
Akzentuierung	0.22
grammatischer Status des Wortes	0.15
Position in der Silbe	0.07
Position des Segments im Wort	0.02

Berndeutsch	BE
Effekt	Verbesserung
Lautklassen	55.49
folgendes Segment	3.92
vorangehendes Segment	1.11
Position der Silbe in der Phrase	0.86
Position in der Silbe	0.68
phonologische Länge des Nukleus	0.48
Anzahl C zwischen V	0.3
Position des Segments im Wort	0.24
übernächstes Segment	0.21
grammatischer Status des Wortes	0.11
Akzentuierung	0.05

Abbildung 7: Faktoren in den Modellen für das Zürichdeutsche und Berndeutsche nach Gewichtung sortiert

Damit wird die unterschiedliche Strukturierung des Daueraspekts deutlich. Aller Wahrscheinlichkeit nach – wenn das größere Corpus nicht der einzige Faktor für die Verbesserung der Daten darstellt – hat der Zürcher Sprecher eine geringere Variation als der Berner Sprecher, was sich als bessere Modellierung äußert. Wenn die Segmentlänge in Bezug auf die übergeordneten Strukturen betrachtet wird, so kann aus dem Vergleich geschlossen werden, dass der Zürcher Sprecher auf der Segmentebene konsistenter ist, dass also gleiche Segmentfolgen unabhängig von deren Position im Wort und in der Silbe eher gleich realisiert werden. Der Berner Sprecher dagegen zeigt eher Modifizierungen durch die Position, d. h. dass sich die Konsistenz eher in übergeordneten Strukturen wie Wörtern und Silben äußert.

7 Mängel des Modells

Die gesamte Modellbildung in Abschnitt 5 zeigt eine schrittweise Annäherung an eine perfekte Verteilung, in der die realen Werte und die durch das Modell errechneten Werte auf einer Linie liegen und die Fehler 0 entsprechen. Diese perfekte Verteilung ist nicht erreicht worden und wird mit realen Daten wohl auch nie erreicht werden können. Rund ein Drittel der Gesamtvarianz ist mit den berücksichtigten Faktoren noch nicht erklärt. Weitere Faktoren können eine Annäherung des Modells an die gemessenen Werte bringen. Dabei ist das Hinzufügen weiterer Faktoren theorie- oder zumindest hypothesengeleitet, was hier exemplarisch dargestellt werden soll.
Die Verteilung der Fehler in Abbildung 6 zeigt, dass das Modell bei den kurzen Segmenten tendenziell eher zu lange Dauerwerte errechnet und bei den langen Segmenten eher zu kurze Dauerwerte, während mittlere Dauerwerte besser erklärt sind. Für die weitere Arbeit müssen demnach eher Faktoren einbezogen werden, die besonders diese Kürzungen von kurzen Segmenten und Dehnungen von langen Segmenten in den realen Daten erklären können. Es ist zu vermuten, dass ein Faktor Sprechgeschwindigkeit eine Verbesserung des Modells hervorbringen kann.
Eine Änderung der Sprechgeschwindigkeit hat nicht lineare Änderungen zur Folge, was im Vergleich von langsamer und schneller französischer Lesesprache von B. Zellner (1998) differenziert aufgezeigt worden ist. Generell verändert eine Erhöhung des Tempos die Phrasierung insofern, als die Phrasen länger werden, zudem sinkt die Pausenfrequenz und deren Dauer wird reduziert. Im Vergleich mehrerer europäischer Sprachen von A. Monaghan (2001) wird deutlich, dass diese Aspekte einerseits übereinzelsprachlich Gültigkeit haben, andererseits in der konkreten Ausgestaltung individuelle Züge tragen. B. Zellner (1998) zeigt zudem auf, dass diese Veränderungen nicht nur auf der Makroebene stattfinden, sondern dass die Neustrukturierung weitere Ebenen trifft: Die Silbifizierung ändert sich, im Französischen erscheinen bei langsamem Tempo zusätzliche Schwa-Silben, Assimilationserscheinungen häufen sich bei schnellerem Sprechtempo. Auf Segmentebene verändert sich insbesondere die intrinsische Dauer der Segmente nicht linear, sondern es liegt für verschiedene Tempi eine je

neue Strukturierung vor[9]. Für die hier vorgestellten Timingmodelle spielen Veränderungen in der Phrasierung, der Silbifizierung und Assimilationen weniger eine Rolle, weil sie in der Transkription bereits berücksichtigt sind und somit auf diese Weise in die Modelle einfließen. Eine Änderung der Sprechgeschwindigkeit in der Synthese nur im Hinblick auf eine andere Phrasierung und verlängerte Pausen, wie sie von J. Trouvain (2002) vorgeschlagen und in Perzeptionstests empirisch abgestützt wurde, reicht für eine statistische Modellierung der Segmentdauern also nicht. Wesentlich wäre hier die Berücksichtigung des Sprechtempos auf die intrinsische Dauer. Bei den vorliegenden spontansprachlichen Daten ist die Sprechgeschwindigkeit jedoch nicht kontrolliert geändert, sondern von den Sprechern ausgehend von einer individuell unterschiedlichen 'Normalgeschwindigkeit' in der Kommunikationssituation spontan und kontinuierlich angepasst worden. Eine detaillierte Analyse der Modellfehler im Bezug auf die einzelnen Segmente, Silben oder Wörter kann hier Hinweise auf eine neue Strukturierung geben. Die Zuordnung einer Sprechgeschwindigkeit zu einzelnen Einheiten ist aber nicht trivial (siehe dazu beispielsweise A. Baltiner et al. 1997). Für die Einarbeitung eines solchen Faktors ist die Basis in unseren Daten noch nicht gegeben, sondern muss zuerst erarbeitet werden.

8 Zusammenfassung und Schlussfolgerungen

Es wurde gezeigt, wie Konzepte der Sprachsynthese für die Linguistik fruchtbar gemacht werden können. Zudem erzwingt die Synthese Bereiche anzusehen, die in rein analytischen Zugängen außer Acht gelassen werden können. Hier wurde die zeitliche Strukturierung angesehen, die in der Prosodieforschung zur Zeit eher nebensächlich behandelt wird. Sie ist für eine Synthese jedoch unumgänglich.
Da in der Konstruktion einer Sprachsynthese für die Mundart nicht von der sonst üblichen Lesesprache ausgegangen werden kann, wurden die Bedingungen für die Mundartsynthese, insbesondere der phonetische Input und die Wahl von Interviews als zu analysierendes Datenmaterial, präsentiert; das ist ein Zugang, der für die Synthese weitgehend neu ist. Im Hauptteil wurde die Bildung eines Timingmodells mittels eines 'generalizded linear models' (GLM) dargestellt. Im Vergleich der Modelle für das Berndeutsche und für das Zürichdeutsche wird die unterschiedliche zeitliche Strukturierung durch die beiden Sprecher unterhalb der Phrasierungsebene deutlich. Der Zürcher Sprecher ist auf der Timing-Ebene monotoner, was sich in der besseren Modellierbarkeit äußert. Die Segmentdauer wird bei ihm stärker durch die segmentalen Aspekte, die direkte Umgebung und die Wortbetonung bestimmt, während beim Berner Sprecher die Position des Segments in der Silbe, im Wort und in der Phrase bedeutsamer ist. Zum Schluss wurde das Problem eines zusätzlichen Faktors 'Sprechgeschwindigkeit' zur Verbesserung des Modells diskutiert.
Diese Resultate zeigen also die unterschiedlichen Timing-Konzepte der beiden Sprecher. Da nur zwei Sprecher untersucht wurden, lässt sich aus dem Bisherigen noch keine schlüssige Aussage darüber machen, ob die hier dargestellten Unterschiede dialektal oder idiolektal bedingt sind. Für die Dialektologie ist es deshalb notwendig, diese Modelle mit zusätzlichen Sprechern zu stützen, bzw. zu verwerfen. Aus diesem Grund wird ein zusätzliches Modell für einen zweiten Berner Sprecher erstellt, für einen zweiten Zürcher Sprecher ist eines geplant.

[9] Da B. Zellner (1998) nur schnelles und langsames Tempo vergleicht, liegen unterschiedliche Kategorisierungen vor. In der spontanen Sprache wird das Sprechtempo aber kontinuierlich geändert, so dass die Reklassifizierung vermutlich auch nicht kategoriell sondern kontinuierlich erfolgt, was jedoch empirisch noch zu überprüfen ist.

Danksagung

Für Unterstützung und Kritik danke ich den ProjektmitarbeiterInnen Eric Keller, Ingrid Hove und Katrin Häsler. Die Forschung ist finanziert durch den Schweizerischen Nationalfonds.

Literaturhinweise

Auer, Peter / Gilles, Peter / Spiekermann, Helmut (Hg.) (2002): Silbenschnitt und Tonakzente. Tübingen [Linguistische Arbeiten 463].
Batliner, A. / Kießling, A. / Kompe, R. / Niemann, H. / Noth. E. (1997): Tempo and its Change in Spontaneous Speech. In: Proceedings of the European Conference on Speech Communication and Technology. Bd. 2. Rhodes, S. 763–766.
Bierwisch, Manfred (1966): Regeln für die Intonation deutscher Sätze. In: Studia Grammatika 7, 99–201.
Brinckmann, Caren / Trouvain, Jürgen (2003): The Role of Duration Models and Symbolic Representation for Timing in Synthetic Speech. In: International Journal of Speech Technology 6, 21–31.
Clauss, Walter (1927): Die Mundart von Uri. Laut- und Flexionslehre. Frauenfeld. [Beiträge zur schweizerdeutschen Grammatik 17].
Féry, Caroline (1993): German Intonational Patterns. Tübingen [Linguistische Arbeiten 285].
Ham, William H. (2001): Phonetic and Phonological Aspects of Geminate Timing. New York & London [Outstanding Dissertations in Linguistics].
Isatschenko, Alexander / Schädlich, Hans-Joachim (1964): Untersuchungen über die deutsche Satzintonation. Berlin.
Keller, Eric (1994). Fundamentals of Phonetic Science. In: Keller, Eric (Hg.): Fundamentals of Speech Synthesis and Speech Recognition. Chichester. S. 5–21.
Keller, Eric / Zellner, Brigitte / Werner, Stefan / Blanchoud, Nicole (1993): The Prediction of Prosodic Timing: Rules for Final Syllable Lengthening in French. Proceedings, ESCA Workshop on Prosody. Lund, Sweden, S. 212–215.
Keller, Eric / Zellner, Brigitte (1995): A statistical timing model for French. XIIIème Congrès International des Sciences Phonétiques. Bd. 3. Stockholm. S. 302–305.
Keller, Eric / Zellner, Brigitte (1996): A timing model for fast French. York Papers in Linguistics 17, 53–75.
Keller, Eric / Bailly, Gérard / Monaghan, Alex / Terken, Jacques / Huckvale, Mark (Hg.) (2001): Improvements in Speech Synthesis. Chichester.
Mixdorff, Hansjörg (2002): An Integrated Approach to Modeling German Prosody. Dresden [Studientexte zur Sprachkommunikation 25].
Mixdorff, Hansjörg / Jokisch; Oliver (2003): Evaluating the Quality of an Integrated Model of German Prosody. In: International Journal of Speech Technology 6, 45–55.
Möbius, Bernd / van Santen, Jan (1996): Modelling Segmental Duration in German Text-to-Speech Synthesis. In: Proceedings of the International Conference on Spoken Language Processing. Bd. 4. Philadelphia, S. 2395–2398.
Monaghan, Alex (2001): An Auditory Analysis of the Prosody of Fast and Slow Speech Styles in English, Dutch and German. In: Keller, Eric et al. (Hg.), S. 204–217.
Neuber, Baldur (1998): Endsilbendehnungen und Tendenz zur temporalen Symmetrie in deutschen Wörtern – Teil 1: Problemdarstellung und Empirie. In: Bose, Ines / Biege, Angela (Hg.): Theorie und Empirie in der Sprachwissenschaft. Festschrift für Eberhard Stock. Halle (Saale), S. 173–181.

Neuber, Baldur (i. Dr.): Endsilbendehnungen und Tendenz zur temporalen Symmetrie in deutschen Wörtern – Teil 2: Theoretische Interpretation. [Erscheint in: Hallesche Schriften zur Sprechwissenschaft und Phonetik].

Ogden, Richard / Local, John / Carter, Paul (1999): Temporal interpretation in Prosynth, a prosodic speech synthesis system. In: Proceedings of the XIVth ICPhS. Bd. 1. San Francisco, S. 1059–1062.

Paolillo, John (i. Dr.): Log-Linear Modeling and Logistic Regression. In: Bayley, Robert / Preston, Dennis (Hg.): Statistical Methods in Linguistics. Amsterdam.

Pierrehumbert, Janet (1980): The phonology and phonetics of English intonation. Diss. Massachusetts Institut of Technology. Cambridge, Mass.

Schmid, Karl (1915): Die Mundart des Amtes Entlebuch im Kanton Luzern. Frauenfeld [Beiträge zur schweizerdeutschen Grammatik 7].

Schmid, Stephan (i. Dr.): Zur Vokalquantität in der Mundart der Stadt Zürich. In: Linguistik online.

Selting, Margret (1995): Prosodie im Gespräch. Tübingen [Linguistische Arbeiten 329].

Siebenhaar, Beat (i. Dr. a): 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 (i. Dr. b): Berner und Zürcher Prosodie – Ansätze zu einem Vergleich. In: Glaser, Elvira / Ott, Peter / Schwarzenbach, Ruedi: Tagungsband der 14. Arbeitstagung zur alemannischen Dialektologie in Männedorf/Zürich 16. – 18.9.2002

Siebenhaar, Beat / Forst, Martin / Keller, Eric (i. Dr. a): Speech Synthesis of Dialectal Variants as a Method for Research on Prosody. In: Proceedings of 'Methods in Dialectology' XI Joensuu (SF).

Siebenhaar, Beat / Forst, Martin / Keller, Eric (i. Dr. b): Timing in Bernese and Zurich German. What the Development of a Dialectal Speech Synthesis System Tells us about it. In: Gilles, Peter / Peters, Jörg (Hg.): Regional Variation in Intonation. Tübingen [Linguistische Arbeiten].

Siebenhaar, Beat / Zellner Keller, Brigitte / Keller, Eric (2001): Phonetic and Timing Considerations in a Swiss High German TTS System. In: Keller, Eric et al. (Hg.), S. 165–175.

Spiekermann, Helmut (2000): Silbenschnitt in deutschen Dialekten. Tübingen [Linguistische Arbeiten 425].

Trouvain, Jürgen (2002): Temposteuerung in der Sprachsynthese durch prosodische Phrasierung. In: Tagungsband 13. Konferenz Elektronische Sprachsignalverarbeitung (ESSV) 2002. Dresden, S. 294–301.

Uhmann, Susanne (1991): Fokusphonologie. Eine Analyse deutscher Intonationskonturen im Rahmen der nicht-linearen Phonologie. Tübingen [Linguistische Arbeiten 252].

van Santen, Jan (1998): Timing. In: Sproat, Richard (Hg.): Multilingual Text-to-Speech Synthesis: The Bell Labs Approach. Dordrecht / Boston / London, S. 115–139.

Vetsch, Jakob (1910): Die Laute der Appenzeller Mundarten. Frauenfeld [Beiträge zur schweizerdeutschen Grammatik 1].

von Essen, Otto (1964): Grundzüge der hochdeutschen Satzintonation. Ratingen.

Weber, Albert (1948): Zürichdeutsche Grammatik. Ein Wegweiser zur guten Mundart. Unter Mitwirkung von Eugen Dieth. Zürich [Grammatiken und Wörterbücher des Schweizerdeutschen in allgemeinverständlicher Darstellung 1].

Willi, Urs (1996): Die segmentale Dauer als phonetischer Parameter von "fortis" und "lenis" bei Plosiven im Zürichdeutschen. Eine akustische und perzeptorische Untersuchung. Stuttgart [Zeitschrift für Dialektologie und Linguistik. Beihefte 92].

Wipf, Elisa (1910): Die Mundart von Visperterminen im Wallis. Frauenfeld [Beiträge zur schweizerdeutschen Grammatik 2].

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. http://www.unil.ch/imm/docs/LAIP/pdf.files/Zellner_Dissertation.pdf

Zellner Keller, Brigitte (2002): Revisiting the Status of Speech Rhythm. In: Bel, Bernard / Marlien, Isabelle (Hg.): Proceedings of the Speech Prosody 2002 Conference, 11-13 April 2002. Aix-en-Provence. S. 727-730.