Phonologie-Phonetik-Kopplung in einem constraintbasierten gesturalen Modell

Markus Walther[1] und Bernd J. Kröger[2]

[1] Seminar für Allgemeine Sprachwissenschaft, Heinrich-Heine-Universität Düsseldorf
Universitätsstr. 1, 40225 Düsseldorf
[2] Institut für Phonetik der Universität zu Köln
Greinstr.2, 50939 Köln

email: walther@sapir.ling.uni-duesseldorf.de, amp04@rs1.rrz.uni-koeln.de

Zusammenfassung Es wird ein implementierter Ansatz zur Kopplung einer constraintbasierten Phonologiekomponente mit einem artikulatorischen Synthesizer vorgestellt. Das Konzept der artikulatorischen Geste schafft dabei eine enge Verbindung zwischen beiden Komponenten, da Gesten physikalisch-phonetische und phonologische Aspekte vereinen. Die Modellierung von Silbenbildung und phonologischen Prozessen wie der Auslautverhärtung im Deutschen geschieht in der Constraintlogikprogrammiersprache CUF. Deren Erweiterung um arithmetische Constraints erlaubt die gleichzeitige Beschreibung von Phonologie und Phonetik und gestattet somit die Anreicherung deklarativer lexikalistischer Grammatiktheorien wie etwa HPSG bis zur Detailebene der phonetischen Realisierung. Erste akustische Demonstrationen zeigen die prinzipielle Tauglichkeit des Ansatzes zur linguistisch fundierten Synthese vollständiger Äußerungen.

An implemented approach which couples a constraint-based phonology component with an articulatory speech synthesizer is proposed. The concept of articulatory gesture ensures a tight connection between both components, as gestures comprise both physical-phonetic and phonological aspects. The phonological modelling of e.g. syllabification and phonological processes such as German final devoicing is expressed in the constraint logic programming language CUF. Extending CUF by arithmetic constraints allows the simultaneous description of both phonology and phonetics. Thus declarative lexicalist theories of grammar such as HPSG may be enriched up to the level of detailed phonetic realisation. Initial acoustic demonstrations show that our approach is in principle capable of synthesizing full utterances in a linguistically motivated fashion.

Markus Walther dankt der DFG im Rahmen des SFB 282 "Theorie des Lexikons", Bernd Kröger u.a. der EG im Rahmen des ESPRIT-Projekts BR 6975 (SPEECHMAPS) für die freundliche Unterstützung. Dank auch an Petra Barg, Jim Kilbury, Tim Skellet und Richard Wiese für die Durchsicht des Manuskripts sowie zwei anonymen Gutachtern für wertvolle Hinweise.

1 Einleitung

In den letzten Jahren ist im Bereich der maschinellen Spracherkennung und –synthese verstärkt die Notwendigkeit zur Berücksichtigung genuin linguistischer, insbesondere phonologischer Information betont worden. Es wird dabei erwartet, daß Robustheit, Erkennungsraten, Natürlichkeit der Aussprache etc. von der Integration linguistischer Erkenntnisse entscheidend profitieren können. Umgekehrt betont in der Phonologie die 'laboratory phonology'-Richtung die Notwendigkeit detaillierter phonetischer Befunde zur Stützung phonologischer Theorien (Kingston 1990). Obwohl also den Fragen der phonetischen Realisierung abstrakter phonologischer Information verstärkte Aufmerksamkeit gilt, gibt es noch kaum praktisch realisierte Modelle, die beide Domänen koppeln und gleichzeitig linguistisch befriedigend sind. Dieses Papier beschreibt einen solchen Ansatz zur Realisierung eines Sprachproduktionsmodells.

Ein wesentlicher Punkt dieses Modells ist, daß es eine computerimplementierte constraintbasierte Phonologiekonzeption verwendet, die die Einbettung in deklarative formale Theorien der Gesamtgrammatik wie etwa HPSG (Pollard & Sag 1987) ermöglicht. Damit ist dann prinzipiell eine simultane Beschreibung des Zusammenwirkens verschiedener linguistisch relevanter Dimensionen (etwa morphologischer Wortakzent, semantischer Fokus, syntaktischer Satzmodus) im Blick auf die phonetische Detailrealisierung gegeben.

Charakteristikum dieses Ansatzes ist, daß keine Schnittstelle zwischen phonologischer und phonetischer *Beschreibung* existiert. Das resultiert aus dem bivalenten Charakter der *artikulatorischen Geste* als grundlegendem Konzept: Sie ist einerseits als abstrakte, zugrundeliegende phonologische Einheit anzusehen.[3] Andererseits aber wird jede Geste als zielgerichtete Bewegung eines bestimmten Artikulators (etwa zur Erreichung eines Lippenverschlusses) realisiert und hat eine konkret definierte zeitliche Ausdehnung und Relation zu anderen Gesten. Im hier beschriebenen Ansatz werden beispielsweise die einer Äußerung zugrundeliegenden zeitlichen Relationen zwischen Gesten in derselben formalen Beschreibung spezifiziert wie etwa phonologische Alternationen.

Die solcherart konkretisierten Gesten bilden unmittelbar die Eingabe für eine Komponente zur artikulatorischer Sprachsynthese. Deren Ziel ist es, in möglichst natürlicher Weise den menschlichen Sprachproduktionsmechanismus zu modellieren. Das verwendete phonetische Produktionsmodell umfasst ein dynamisches Modell zur Generierung von Artikulatorbewegungen auf der Basis quantitativ-temporal vollspezifizierter Gesten und ein artikulatorisch-akustisches Modell, d.h. einen künstlichen Sprechtrakt. Dieser sorgt für die Generierung des akustischen Sprachsignals, das letztendlich die abstrakte linguistische Ausgangsinformation hörbar macht. Durch diese Rückkopplung an die empirische Wirklichkeit eröffnet sich ein großes Potential zur Validierung linguistischer, insbesondere phonologischer Theoriebildung, da deren Einzelaspekte im Modell kontrollierbar

[3] Gesten entsprechen *nicht* der traditionellen Kategorie 'Phonem'. Die Verhältnisbestimmung ist problematisch – in grober Näherung könnte man einem Phonem eine oder mehrere Gesten zuzuordnen.

sind und beispielsweise Perzeptionstests unterworfen werden können.

2 Das dynamische Modell und die Geste

Die Kinematik der Artikulatorbewegungen wird mittels des dynamischen Modells aus den Gesten berechnet.

Phonetisch entsprechen Gesten räumlich zielgerichteten und zeitlich lokalisierbaren Artikulatorbewegungen, wobei die Intention die Ausbildung einer linguistisch relevanten Konstriktion ist (Saltzman und Munhall 1989; Browman & Goldstein 1992). Das Ziel wird als konkreter Targetpunkt im Steuerparameterraum des entsprechenden Artikulators quantifiziert. Jede Geste wirkt nur innerhalb eines definierten Zeitintervalls, dem gesturalen Aktivintervall auf den die Geste realisierenden Artikulator. Wenn für einen Artikulator keine Geste aktiv ist, so wird die artikulatorinhärente Neutralposition angestrebt. Die Neutralposition aller Artikulatoren entspricht im Modell der Produktion eines stimmhaften nichtnasalen Schwas. Physikalisch wird die Bewegung eines gestural aktivierten Artikulators mittels des kritisch gedämpften harmonischen Oszillators beschrieben.[4] Es kann für jede Geste eine inhärente Zeitskala, die Phasenskala definiert werden. Phasenwerte parametrisieren den relativen Artikulator- Target-Abstand, d.h. den Realisierungsgrad der Geste: Bei einem bestimmten Phasenwert liegt immer der gleiche relative Artikulator-Target-Abstand vor (Kröger 1993a, 1993b). Der inhärente Gestenparameter *Eigenperiodendauer* gibt die absolute Länge des Zeitintervalls von 0 bis 360 Grad an, definiert somit die zeitliche Dehnung der Phasenskala einer Geste. Eigenperiodendauern unterscheiden sich für vokalische, konsonantische und Öffnungsgesten (Glottis- und Velumsgesten). Die Herabsetzung der Eigenperiodendauer entspricht der schnellen Ausführung der Geste: Ein konstanter relativer Artikulator-Target-Abstand wird dann schneller erreicht.

Die zeitliche intergesturale Koordinierung aller zur Realisierung einer Äußerung notwendigen Gesten wird in relativer Form von Geste zu Geste durchgeführt. Dazu wird der Gestenparameter *Assoziationsphase* benötigt. Der dem Assoziationsphasenwert der einen Geste entsprechende abstrakte Zeitpunkt wird einem ebensolchen Phasenpunkt der anderen Geste gleichgesetzt.[5] Eine detaillierte Beschreibung dieser Phasenregeln wird von Kröger (1993b) gegeben. Grundsätzlich werden dabei zunächst die vokalischen Gesten ohne Ausbildung einer zeitlichen Lücke hintereinandergefügt. Die Zeitpunkte des Anfangs bzw. Endes eines vokalischen Aktivintervalls stellen nun die Ausgangspunkte für die Koordinierung mit den konsonantischen Schließgesten dar. Somit entsteht immer eine zeitliche Überlappung von konsonantischen und vokalischen Gesten und die resultierende Koproduktion modelliert automatisch Aspekte der Koartikulation.

[4] Zur Veranschaulichung stelle man sich eine an einer Feder hängende Masse vor, die in einem hochviskosen Medium aus der Ruhelage ausgelenkt wird und zu schwingen versucht.

[5] Siehe etwa die Berechnung des Endzeitpunkts einer Geste in (6).

Der Gestenparameter *Ablösephase* schließlich bestimmt das abstrakte Ende des Aktivintervalls und ist wie die Assoziationsphase kontextabhängig.

3 Das artikulatorisch–akustische Modell

Das Artikulationsmodell der artikulatorisch-akustischen Komponente berechnet anhand der vorgegeben Artikulatorbewegungen die Gestalt der Längsschnittfläche (Mittsagittalschnitt) und daraus die dreidimensionale Geometrie des Ansatzrohres. Zusätzlich berechnet es aus den Steuerparametern *Stimmlippenspannung* und *glottaler Öffnungsgrad* (d.h. mittlerer Stimmlippenabstand, artikulatorische Komponente des Stimmlippenabstandes) die momentane glottale Öffnungsfläche. Anhand dieser geometrischen Daten kann dann mittels des akustischen Modells das Sprachsignal berechnet werden. Das Artikulationsmodell definiert die Modellartikulatoren und den Wertebereich der zugehörigen Steuerparameter. Da die Akustik des Sprachsignals neben der phonatorischen Einstellung allein von der Geometrie des Ansatzrohrs bestimmt wird, wurden die Steuerparameter derart definiert, daß sie direkt Konstriktionsgrad und –ort des zugehörigen Artikulators beschreiben.

Das akustische Modell umfaßt ein Sprechtraktmodell und ein selbstschwingendes Glottismodell. Das Sprechtraktmodell berechnet die im Sprechtrakt vorliegenden Druck- und Luftstromverhältnisse an jedem Ort im Sprechtrakt zu jedem Zeitpunkt. Friktionsrauschen wird automatisch generiert, wenn die aerodynamischen Bedingungen für das Entstehen turbulenter Strömungen erfüllt sind.[6] Das selbstschwingende Glottismodell wird mittels der Parameter *Stimmlippenspannung* und *Glottisöffnungsgrad* gesteuert, wobei letztere die artikulatorisch aktiv beeinflußbare Einstellung der Stimmlippen bezeichnet. Deren Schwingung entsteht somit nur bei geeigneter Spannung der Stimmlippen zusammen mit passendem subglottalen Druck.

4 Die phonologische Komponente

Die phonologische Komponente ist für die Modellierung der sprachspezifischen und nicht physikalisch–biologisch determinierten Anteile des lautlichen Subsystems zuständig. Wir legen dabei das Paradigma der Deklarativen Phonologie (DP, cf. Scobbie 1991) zugrunde, in welchem im Gegensatz zum bisherigen generativen Ansatz mit seiner prozeduralen Ausrichtung nur Wohlgeformtheitsbedingungen für eine Lösung konstatiert werden, aber nicht die Schritte zur Lösungsfindung (Deklarativität). In DP wird eine extrinsische Regelordnung als überflüssig angesehen, phonologische Regeln und Repräsentationen sind formal gleichwertig, Merkmalswertänderungen sind unmöglich (Monotonizität) und es wird auf Ableitungsebenen verzichtet (Monostratalität). Phonologische Beschreibungen stehen ihrem Denotat in der realen Welt, Klassen von Schallereignissen,

[6] D.h. genügend enge Konstriktion im Sprechtrakt bei gleichzeitig genügend hohem Luftstrom.

gegenüber. Mit Bird (1991:143) verlangen wir, daß phonologische Beschreibungen darüberhinaus formal adäquat sein müssen, d.h. in einer Beschreibungssprache mit formaler Syntax und Semantik auszudrücken sind. Hier wird dazu CUF (Dörre & Eisele 1991) herangezogen. CUF ist eine für linguistische Anwendungen konzipierte Constraintlogikprogrammiersprache,die etwa die Implementierung von HPSG erlaubt. Vorarbeiten zur deklarativen Silbifizierung und Formalisierung phonologischer Prozesse sind in (Walther 1992, 1993 ; Walther & Wiese 1993) dokumentiert.

Hier soll nun am Beispiel des Wortes **ebbt** *(ab)ebben, 3sg pres.ind.* das generelle Vorgehen demonstriert werden. Wie **ebbt** – [ʔɛpt^h] vs. **ebben** – [ʔɛbən] zeigt, alterniert der labiale Plosiv kontextabhängig bezüglich Stimmhaftigkeit ([p] vs. [b]). Dies ist eine Instanz des als Auslautverhärtung bekannten Prozesses im Deutschen, der durch die folgenden zwei in der Literatur kaum noch umstrittenen oberflächentreuen Generalisierungen beschrieben werden kann (Beispiele jeweils unter b.):

(1) a. Obstruenten sind in der Silbenkoda (`coda`) stimmlos.

b. Jagd [jaːkt.], Bund [bUnt.], Lob [loːp.], ebbt [ɛpt.], brav [bʁaːf.]

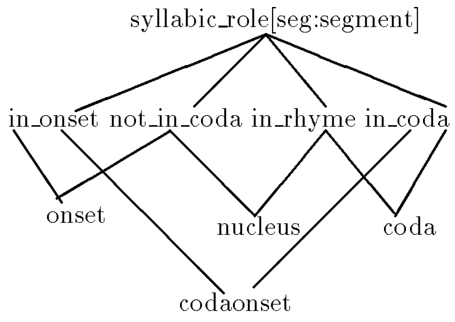
(2) a. Alternierende Obstruenten sind im Silbenansatz (`in_onset`) stimmhaft.

b. jagen [jaː.gən], Bundes [bUn.dəs], Lobes [loː.bəs], Ebbe [ɛb̄ə], braver [bʁaː.vɐ]

Im folgenden stellen wir die Schritte zur formalen Modellierung am Beispiel dar. Zunächst werden Segmente (`seg`) als aus primären und sekundären Gesten bestehende Einheiten spezifiziert. In CUF: `segment :: primary:gesture, secondary:gesture`. Phonologisch kann die Geste als zugrundeliegende abstrakte Einheit aufgefasst werden (Browman & Goldstein 1989, 1992). Hier definiert sie diskrete Kategorien wie z.B. Artikulator, Artikulationsort und den Verengungsgrad. Kategorien für Artikulatoren oder Artikulationsorte sind beispielsweise *Lippen, Zungenspitze, Stimmlippen*, oder *alveolar, postalveolar, palatal, velar*. Kategorien für Artikulationsart sind beispielsweise *Vollverschluß-*, *Beihnahverschluß-* oder *Näherungsgeste*, zur Realisierung eines Plosivs, Frikativs oder einer vokalischen Konstriktion.

Als nächstes müssen segmentale Positionen mit Silbenstruktur verbunden werden. Dazu wird ein entsprechender Typverband, wie aus der HPSG (Pollard & Sag 1987) bekannt, bereitgestellt (3).

(3)



Durch die in (3) erfolgte Dekomposition subsilbischer Rollen wird etwa ein Silbenkern (`nucleus`) als gleichzeitig im Reim und nicht in der Koda der Silbe repräsentiert. Ambisilbische Positionen (`codaonset`), wie sie etwa für **Ebbe** benötigt werden, können bemerkenswerterweise durch die statische Typbeschreibung allein gekennzeichnet werden und müssen nicht durch konfigurationelle Dominanz beschrieben werden.

Die Generalisierungen in (1) und (2), die im generativen Ansatz üblicherweise durch eine das *Merkmal Stimmhaftigkeit ändernde Regel* über zugrundeliegend stimmhaften Repräsentationen der alternierenden Obstruenten beschrieben wurden, werden im folgenden als monotone konditionale Constraints umformuliert. Die prädikatenlogische und CUF-Formalisierung sind dabei jeweils gegenübergestellt. Letztere basiert auf bekannten logischen Umformungen unter Verwendung der logischen Konnektive Negation (~), Disjunktion (;) und Konjunktion (&) zur Verknüpfung der Attribut:Wert-Beschreibungen.

(4) a. $\forall\, segment\ S : obstruent(S) \rightarrow (coda(S) \rightarrow stimmlos(S))$
   b. `final_devoicing := self:(seg:~obstruent ; seg:obstruent & (~coda ; coda & seg:secondary:voiceless)).`

(5) a. $in\_onset(S) \rightarrow stimmhaft(S)$
   b. `voiced_in_onset := self:(~in_onset ; in_onset & seg:secondary:inactive).`

(4) kann als lexikalisches Constraint über dem segmentalen Inventar aufgefaßt werden, während (5) Bestandteil angereicherter segmentaler Repräsentationen für die Klasse der alternierenden Obstruenten (/b,d,g,v,z,ʒ/) bildet. In Übereinstimmung mit dem lexikalistischen Ansatz der HPSG sind diese Auslautverhärtungs*repräsentationen* also Bestandteil normaler Lexikoneinträge.

Wie kommt nun die Phonetik in die abstrakte phonologische Repräsentation? Hier wird davon ausgegangen, daß *phonologische Beschreibung unterspezifizierte phonetische Beschreibung* ist. Wir fügen dementsprechend einfach zusätzliche Informationen hinzu, die vor allem die quantitativen Dimensionen der Geste reflektieren: `gesture :: cd:constriction_degree, cl:constriction_location, clip`[7]`:clipping, time:timing. timing :: start:number, end:number,`

[7] Zu Clipping siehe (Kröger 1993a).

eigenperiod:number, assoc:number, release:number. Neben dem vorher schon vorhandenen Engegrad und Ort der Engebildung kommen insbesondere die Eigenperiodendauer und Assoziations- und Ablösephase hinzu.[8] Diese Steuerparameter werden für die einzelnen Gesten mit konkreten Werten gefüllt (aus Kröger 1993a,b). Um nun die vom Silbenkontext abhängige und nicht lexikalisch spezifizierte absolute zeitliche Lage der Gesten für eine gegebene Äußerung zu instantiieren, werden die Assoziationsregeln von Kröger (1993b) zur Fixierung des intergesturalen Timings in geeignet angepaßter Form[9] in CUF durch einfache rekursive Definitionen formuliert. Zur grundsätzlichen Behandlung der Zeitdomäne ist dazu eine Erweiterung von CUF um *arithmetische Constraints* nötig. Sie gestattet es, zusätzlich zu den vorhandenen Featureterm-Constraints Gleichungen mit den üblichen arithmetischen Operationen über der Domäne der reellen Zahlen zu formulieren und diese auch praktisch zu lösen. Damit kann etwa die Berechnung des Endzeitpunkts einer Geste durch $endtime_{Geste} = start_{Geste} + eigenperiod * release_{Geste}/360$ auch bei zum Berechnungszeitpunkt *unbekanntem* Startwert erfolgen, was zur Erhaltung des ordnungsfreien Charakters von CUF wesentlich ist. In CUF ist dies wie in (6) ersichtlich formulierbar.[10] Anschließend folgt dort die volle Segmentdefinition für /**b**/. Mit den globalen Constraints zur Silbifizierung, Assoziationsfestlegung und Abgrenzung von phonologischen Wörtern kann dann schließlich das Beispiel wie am Ende von (6) ersichtlich repräsentiert werden.

(6)
```
endtime := time:(start:Start & end:End & eigenperiod:EigenPeriod &
                 release:Release) &
           equal(phase_point(Start, EigenPeriod, Release), End).
phase_point(Start, EigenPeriod, Phase) := add(Start,
     multiplicate(EigenPeriod, divide(Phase, 360))).

b := is(supralaryngeal_obstruent)
     & primary_articulation(lips & closed(thelips) & endtime
     & association_and_release_phase(stop)
     & obstruent_clipping(default))
     & voiced_in_onset & final_devoicing & theeigenperiod.
global_constraints := syllabify & is_phonological_word
     & association_rules.
ebbt := phon:[glottal_stop, ae, b, t, postphonatory_opening].
```

[8] Start- und Endzeitpunkt des gesturalen Aktivintervalls sind *keine* gesturalen Steuerparameter, sondern in der endpunktbasierten Charakterisierung des Zeitintervalls begründet.

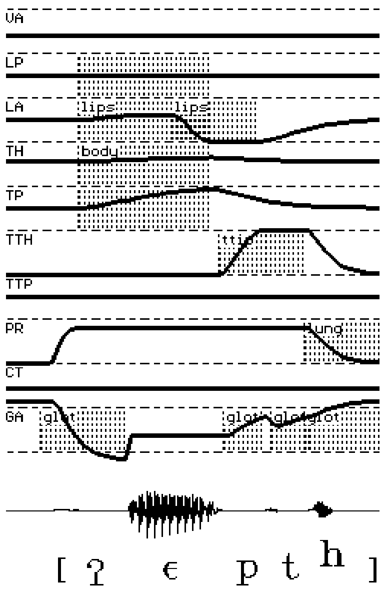
[9] Referenz auf subsilbische Konstituenten statt auf konsonantische vs. vokalische Gesten.

[10] Die zur Fertigstellung diese Papiers vorliegende Implementierung simuliert arithmetische Constraints noch durch in Featuretermen ausgedrückte arithm. Formeln, die einem um arithm. Constraints erweiterten PROLOG zur Gleichungslösung übergeben werden. Wie jedoch (Walther 1992) an einem Forschungsprototyp von CUF bereits gezeigt hat, ist die volle Integration mithilfe eines entsprechenden PROLOG-Derivats (CLP($\mathcal{R}$) oder DMCAI-Sicstus-PROLOG-Clone) unproblematisch und für die aktuelle CUF-Version in Vorbereitung.

```
test(beispiel) := ebbt & phon:global_constraints.
```

Aus **test(beispiel)** ermittelt der CUF-Constraintlösungsmechanismus eine komplexe Struktur, von der (7) den Ausschnitt für **/b/[p]** zeigt:

$$
(7)\quad
\begin{bmatrix}
\textit{coda} \\
\text{seg}\;
\begin{bmatrix}
\textit{supralaryngeal\_obstruent} \\
\text{primary}\;
\begin{bmatrix}
\textit{lips} \\
\text{cd}\;\begin{bmatrix}\textit{constriction\_degree}\\ \text{art}\;\;2 \\ \text{val}\;-20\end{bmatrix} \\
\text{clip}\;\begin{bmatrix}\textit{obs}\\ \text{val}\;0\end{bmatrix} \\
\text{time}\;\begin{bmatrix}\textit{timing}\\ \text{start}\quad 178.889 \\ \text{end}\quad 267.778 \\ \text{eigenperiod}\;80 \\ \text{assoc}\quad 180 \\ \text{release}\;400\end{bmatrix}
\end{bmatrix} \\
\text{secondary}\;
\begin{bmatrix}
\textit{glottis} \\
\text{cd}\;\begin{bmatrix}\textit{constriction\_degree}\\ \text{art}\;\;10 \\ \text{val}\;400\end{bmatrix} \\
\text{clip}\;\begin{bmatrix}\textit{null}\\ \text{val}\;0\end{bmatrix} \\
\text{time}\;\begin{bmatrix}\textit{timing}\\ \text{start}\quad 232.889 \\ \text{end}\quad 272.889 \\ \text{eigenperiod}\;80 \\ \text{assoc}\quad 157 \\ \text{release}\;180\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$



Diese getypte Attribut-Werte-Struktur enthält *alle* Information, die der artikulatorische Synthesizer zur Berechnung des von ihr denotierten Schallereignisses benötigt.[11] Insbesondere dokumentiert sich die fehlende explizite Schnittstelle in der Abwesenheit etwa eines zusätzlichen `phonetics`-Attributes mit entsprechender komplexer Interpretationsbeziehung zwischen `phon`- und `phonetics`-Werten. Eine Visualisierung der gesturalen Trajektorien und Zeitbeziehungen sowie das Oszillogram von **ebbt** zeigt (8).

5 Diskussion und Zusammenfassung

Die in (Bird 1992, Bird & Klein 1993) beschriebenen finite-state bzw. unifikationsbasierten Ansätze propagieren ebenfalls explizit die Anreicherung von HPSG um deklarativ spezifizierte phonologische Information, sind jedoch nicht quantitativ-phonetisch ausgelegt.

[11] Aus technischen Gründen wird noch eine triviale Formatkonversion mit Extraktion der phonetikrelevanten Attribute durchgeführt.

YORKTALK (Coleman 1992) ist dagegen ein explizit auf phonetische Realisierung ausgelegtes Synthesemodell im Rahmen des DP-Paradigmas. Nichtsegmentale phonologische Beschreibungen und sehr umfangreiche 'phonetic exponency'-Spezifikationen[12] dienen zur Steuerung eines Klatt-Synthesizers auf der *akustischen* Ebene. Gesten als elegante bivalente Primitive operieren hingegen auf der natürlicher erscheinenden *artikulatorischen* Ebene und vereinfachen unserer Ansicht nach die phonetische Realisierung aus der Sicht der Phonologie deutlich.

Auch das SYNPHONICS Concept-to-Speech-System strebt die phonetische Interpretation im Rahmen der HPSG an. Unter anderem wird in (Günther 1993) eine interessante direkte Verbindung von semantischem Fokus zur phonetischen Akzentrealisierung diskutiert. Durch die projektierte Verwendung des Klatt-Synthesizers teilt dieser Ansatz jedoch Probleme von YORKTALK und dupliziert darüberhinaus durch die Trennung von phonologischer und phonetischer Domäne in den Attribut-Werte-Strukturen offenbar Information. Die genaue Vermittlung zwischen ersterer und letzterer, ebenfalls gestural aufgefaßter Domäne bleibt vorerst unklar.

Der vorgestellte Ansatz zur Phonologie-Phonetik-Kopplung steckt ohne Zweifel erst in den Anfängen. Zum einen muß das Studium der Zeitbeziehungen von Gesten im Zusammenhang mit dem Einfluß der Silbenstruktur vorangetrieben werden. Auch andere linguistische Dimensionen wie Wortakzent etc. sind in ihren Timing-Auswirkungen zu studieren, wobei maschinelles Lernen parametrischer Zeitmodelle aus einem gestural segmentierten und annotierten Korpus einen vielversprechenden Ansatz darstellen dürfte (cf. Santen 1993).

Daneben ist die tatsächliche Realisierung phonologischer Prozesse zu untersuchen und zu modellieren. Ist Auslautverhärtung möglicherweise graduell und durch pragmatische Faktoren konditioniert (Port & Crawford 1989)? Ist die ŋ-Realisierung (**Diphthong** [diftɔŋ] – **diphthongieren** [. . . ŋgiːʁən]) als intrasegmentale Variation der gesturalen Phasenbeziehungen auffaßbar? Wie steht es mit den silbenabhängigen Prozessen der r-Allophonie (**frisch** [fʁiʃ] – **Berg** [b ɛʁ̞̯k] – **Bauer** [bauɐ̯]), der ich/ach-Alternation und g-Spirantisierung[13] und der Schwa-Epenthese? Ist letztere möglicherweise in vielen Kontexten als kurze gestural unspezifizierte zeitliche Lücke modellierbar, wie in Walther (1992) vermutet wird? Im Tandem mit elektroartikulographischen Experimenten sind solche Hypothesen sehr konkret in unserem Ansatz testbar.

Der vorgestellte Ansatz zur Phonetik-Phonologie-Kopplung geht von einer deklarativen, constraintbasierten Phonologiekomponente aus, die in der um arithmetische Constraints erweiterten Constraintlogikprogrammiersprache CUF realisiert wurde (Walther 1992, 1993). Zentral für die einfache und schnittstellenfreie Kopplung zum artikulatorischen Synthesizer von Kröger (1993a,b) war das Konzept der Geste. Erste praktische Syntheseergebnisse demonstrieren die

[12] Sie wurden bis auf isolierte Beispiele nicht veröffentlicht, stellen jedoch die Hauptquelle der Qualität von YORKTALK's Sprachausgabe dar (persönliche Mitteilung Richard Ogden).

[13] Cf. Walther & Wiese (1993) zur Formalisierung der letzten beiden Prozesse.

Tragfähigkeit des Ansatzes und öffnen den Weg zu einer empirisch fundierten Abgrenzung zwischen kategoriell-phonologisch und graduell-phonetisch zu beschreibenden Phänomenen relativ zu einem implementierten Gesamtmodell.

Literatur

Bird, S. (1991): Feature Structures and Indices. *Phonology* 8:137–144.

Bird, S. (1992): Finite-State Phonology in HPSG. In: *Proceedings of the 15th International Conference on Computational Linguistics*, Nantes.

Bird, S. & Klein, E. (1993): Enriching HPSG Phonology. Research Paper EUCCS/RP-56, Centre for Cognitive Science, Edinburgh University.

Browman, C. P. & Goldstein, L. (1989): Articulatory gestures as phonological units. *Phonology*, 6:201–251.

Browman, C.P. & Goldstein, L. (1992): Articulatory phonology: An overview. *Phonetica* 49:155-180.

Coleman, J. (1992): The phonetic interpretation of headed phonological structures containing overlapping constituents. *Phonology* 9:1–44.

Dörre, J. & Eisele, A. (1991): A Comprehensive Unification-Based Grammar Formalism. DYANA Deliverable R3.1.B, Universität Stuttgart.

Günther, C. (1993): Planung und Repräsentation prosodischer und phonetisch-artikulatorischer Merkmale im Sprachproduktionssystem SYNPHONICS. In: Abstractband DGfS/CL'93, 4. Fachtagung Sektion Computerlinguistik der DGfS, Universität Hamburg, 17.-19.11.1993.

Kingston, John & Beckman, M. E. (1990) (ed.): Between the grammar and the physics of speech. Papers in laboratory phonology: 1. Cambridge University Press

Kröger, B.J. (1993a): A gestural production model and its application to reduction in German. *Phonetica* 50, 213-233.

Kröger, B.J. (1993b): A gestural approach for controlling an articulatory speech synthesizer. Proceedings EUROSPEECH 93, 1903-1907, Berlin.

Pollard, C. J. & Sag, I. A. (1987): Information-based Syntax and Semantics. CSLI Lecture Notes 13, Chicago University Press.

Port, R. & Crawford, P. (1989): Incomplete neutralization and pragmatics in German. *Journal of Phonetics*, 17:257–282.

Saltzman, E.L. & Munhall,K.G. (1989): A dynamical approach to gestural patterning in speech production. *Ecological psychology*, Vol.1,pp.333-382.

van Santen, J.P.H. (1993): Timing in Text-to-Speech Systems. Proceedings EUROSPEECH 93, 1397-1406, Berlin.

Scobbie, J. (1991): Towards Declarative Phonology. In: Bird, S. (ed.), *Declarative Perspectives on Phonology*, chapter III, p. 1–26. Centre for Cognitive Science, Edinburgh.

Walther, Markus (1992): Deklarative Silbifizierung in einem constraintbasierten Grammatikformalismus. Diplomarbeit Nr.862, Institut für Informatik, Universität Stuttgart.

Walther, Markus (1993): Declarative Syllabification with Applications to German. In: Ellison, M. & Scobbie, J. (eds.), *Computational Phonology*, Edinburgh Working Papers in Cognitive Science 8.

Walther, Markus & Wiese, Richard (1993): Deklarative vs. prozedurale Modellierung von Konsonantenalternationen im Deutschen. In: Abstractband DGfS/CL'93, Universität Hamburg.