

Zur phonetischen Realisierung von Sprechtempoänderungen unter Einbeziehung von artikulatorischer Reorganisation: Artikulatorische und perzeptive Untersuchungen

Bernd J. Kröger

Increase in speaking rate can be produced by two different articulatory strategies: increase of articulator velocity or increase in temporal overlap of gestures. The second strategy may lead to segmental changes (e.g. assimilation or elision). In current speech synthesis systems these changes can only be generated by explicit segmental rules. In this paper an implemented approach is introduced in which a variety of segmental changes simply result from an increase in temporal overlap of articulatory gestures without changing the underlying gestural structure of the utterance. However, segmental changes require articulatory reorganisation in some cases.

Sprechtempoerhöhung kann durch zwei unterschiedliche artikulatorische Strategien realisiert werden: Erhöhung der Artikulatorgeschwindigkeit oder der zeitlichen Überlappung von Gesten. Letztere kann auch zu segmentalen Änderungen (z.B. Assimilation oder Elision) führen. Diese Änderungen sind in herkömmlichen Synthesesystemen nur über explizite segmentale Regeln generierbar. Hier wird ein implementierter Ansatz vorgestellt, in dem viele dieser segmentalen Änderungen allein aufgrund einer zunehmenden zeitlichen Überlappung von artikulatorischen Gesten ohne Änderung der zugrundeliegenden gestischen Struktur entstehen. Allerdings existieren auch segmentale Änderungen, die nur durch artikulatorische Reorganisationsprozesse realisiert werden können.

18.1 Einleitung

Die Einbeziehung von Sprechtempoänderungen in Sprachsynthesystemen führt einerseits zu einer Steigerung der Natürlichkeit und ist andererseits aus praktischer Sicht – beispielsweise zur Optimierung eines Vorleseautomaten für Blinde – von Wichtigkeit. Jedoch zeigt sich schnell, daß Sprechtempoänderungen nicht einfach durch eine lineare Transformation wie etwa durch proportionale Kürzung aller Lautdauern realisiert werden. Vielmehr treten unterschiedlich starke Lautdaueränderungen in verschiedenen Bereichen einer Äußerung auf, und es kann auch zu diskreten segmentalen Änderungen (Reduktionen, siehe Kohler 1990) kommen. Es wird gezeigt, daß die Modellierung von Sprechtempoänderungen und Reduktion gerade vom gestischen Konzept der artikulatorischen Steuerung (Browman und Goldstein 1990), das von uns als Teil eines implementierten artikulatori-

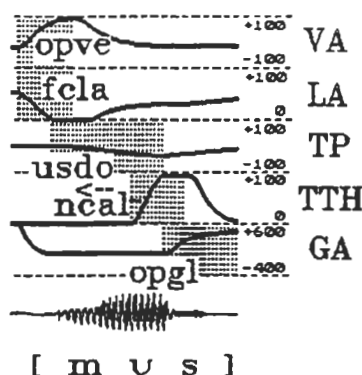


Abbildung 18.1: Steuerparameterzeitfunktionen.

schen Sprachsynthesystems (Kröger 1993) realisiert wurde, in einfacher Weise geleistet werden kann.

18.2 Das gestische Modell der Sprachproduktion

Die Geste kann sowohl als artikulatorische Bewegungseinheit als auch als Grundeinheit der Artikulatorischen Phonologie (Browman und Goldstein 1986, 1989, 1992) angesehen werden. Hierdurch kann die Einführung einer Schnittstelle zwischen phonologischer Darstellung und phonetischer Realisierung einer Äußerung vermieden werden (Walther und Kröger 1994). Liberman und Mattingly (1985) postulieren innerhalb der Motor-Theorie die Geste als grundlegende Einheit der Sprachproduktion wie der Sprachperzeption. Fujimura (1986) etabliert die Geste als invariante artikulatorische Einheit, indem er anhand von artikulatorischen Messungen nachweist, daß artikulatorische Auslenkungs-Zeit-Muster keine stabilen Targetbereiche, d.h. keine stabilen artikulatorischen Maxima und Minima, wohl aber stabile Konturen im Bereich der stärksten Bewegung und damit im Bereich der höchsten Geschwindigkeit eines Artikulators aufweisen. Insbesondere stützt auch die Generierung von segmentalen Änderungen allein aufgrund zunehmender zeitlicher Überlappung der Gesten einer Äußerung die Annahme, daß Gesten als zugrundeliegende Einheiten der Sprachproduktion angesehen werden können: Gesten bleiben unter Sprechtempoänderung erhalten, während die segmentale Struktur der Äußerung variieren kann (Kapitel 18.3.3).

In konkreten Realisierungen des gestischen Konzeptes als implementiertes Modell (Saltzman 1985; Kelso et al. 1986; Saltzman und Munhall 1989; Kröger 1993) wird die Geste als artikulatorische Steuereinheit aufgefaßt, wobei das Ziel jeder einzelnen Geste die Ausbildung einer linguistisch relevanten Sprechtraktkonstriktion ist. Abbildung 18.1¹ zeigt die von unse-

¹Erläuterung zu Abb. 18.1. Dicke Linien für die Steuerparameter Velumsöffnung VA, Lippenöffnung LA, Zungenrückenposition TP, Zungenspitzenhöhe TTH und Glot-

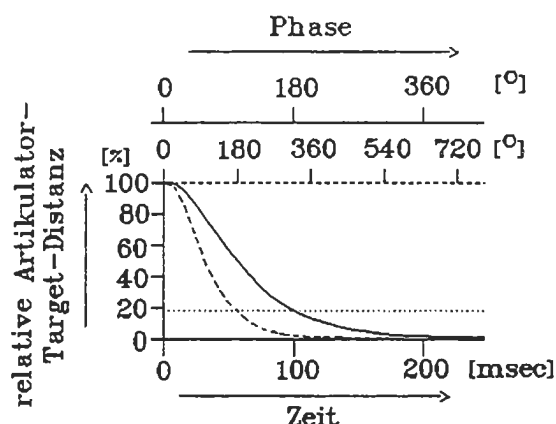


Abbildung 18.2: Zeitfunktion von gestisch gesteuerten Artikulatorbewegungen mit verschwindender initialer Artikulatorgeschwindigkeit.

rem implementierten Produktionsmodell (Kröger 1993) generierten Steuerparameterzeitfunktionen für das Wort [mʊs] und das Oszillogramm des zugehörigen synthetischen Sprachsignals.

Die schattierten Bereiche kennzeichnen die gestischen Aktivitätsintervalle, d.h. die Zeitbereiche, in denen ein Artikulator von einer Geste gesteuert wird. Das gestische Ziel ist durch den Namen der Geste gekennzeichnet (z.B. die Ausbildung eines labialen Verschlusses, einer vokalischen Zungenrückenkonstruktion zur Realisierung des /ʊ/ oder einer glottalen Öffnung).

Ganze Äußerungen werden aus einer Vielzahl von Gesten aufgebaut, die in definierter Weise untereinander zeitlich koordiniert sind. Die gestische Koordinierung wird durch Assoziationslinien innerhalb der „gestischen Partitur“ einer Äußerung gekennzeichnet (siehe z.B. Abbildung 18.3, Kapitel 18.3.3, und Kröger 1993).

Gesten können durch unterschiedliche Parameter qualitativ und quantitativ beschrieben werden. Der qualitative Parameter „ausführender Artikulator“ determiniert den Artikulator, der von einer Geste während ihres Aktivitätsintervalls gesteuert wird (Lippen, Zungenrücken etc.). Die quantitativen gestischen Parameter resultieren aus dem der Geste zugrunde liegenden dynamischen Konzept, dem physikalischen Konzept des kritisch gedämpften harmonischen Oszillators (Saltzman 1985; Kröger 1993). Dieses physikalische Modell führt zu monotonen targetgerichteten Artikulatorbewegungen während eines gestischen Aktivitätsintervalls (Ab-

tisöffnung GA), gestische Aktivitätsintervalle (schattierte Rechtecke) und das generierte Oszillogramm des synthetischen Sprachsignals von [mʊs]. Die Gesten sind mittels vierbuchstabiger Etiketten gekennzeichnet: Öffnungsgeste des Velums (opve), labiale Vollverschlußgeste (fcla), dorsale Vokalgeste zur Realisierung eines ʊ/ (usdo), alveolare Beinahverschlußgeste (ncal) und glottale Öffnungsgeste (opgl).

bildung 18.2). Wichtige quantitative Parameter sind „Targetposition“ TP, „Eigenperiodendauer“ TE, „Ablösephase“ RP (release phase) und „Assoziationsphase“ AP. Die Targetposition determiniert das Ziel der gestischen Bewegung eines Artikulators. Die Eigenperiodendauer beschreibt das von der Geste benötigte Zeitintervall, bis die aktuelle Artikulator-Target-Distanz relativ zur Artikulator-Target-Distanz zu Beginn der gestischen Aktivierung auf 1.4% abgesunken ist. Dies entspricht einem Zeitintervall, das benötigt wird, bis das Target vom Artikulator weitgehend erreicht ist. Eine Geste mit geringerer Eigenperiodendauer erreicht somit eine definierte relative Artikulator-Target-Distanz eher als eine Geste mit höherer Eigenperiodendauer (Abbildung 18.2).² Im Falle vergleichbarer initialer Artikulator-Target-Distanz entspricht eine geringe (hohe) Eigenperiodendauer dann direkt einer hohen (niedrigen) Artikulatorgeschwindigkeit. Aufgrund des Modells des harmonischen Oszillators kann für jede Geste eine intrinsische Zeitskala, die „Phasenskala“, definiert werden. Dabei entspricht das Intervall von 0° bis 360° genau dem durch die Eigenperiodendauer der Geste definierten Zeitintervall (Abbildung 18.2). Im Fall des Verschwindens der initialen Artikulatorgeschwindigkeit ist die gestisch gesteuerte Artikulator-Target-Distanz nur Funktion der Phase. Somit hat die Benutzung der Phasenskala als Zeitskala den Vorteil, daß die auftretenden Phasenwerte im Unterschied zu absoluten Zeitwerten in direkter Weise den relativen Artikulator-Target-Abstand und damit den Grad der (momentanen) Realisierung einer Geste parametrisieren: Ein Phasenwert gibt immer den Grad der Realisierung einer Geste an, d.h. den Grad der erreichten Annäherung zwischen Artikulator und Target. Beispielsweise wird bei Vollverschlußgesten der Verschluß im Ansatzrohr nur dann ausgebildet, wenn die Geste unabhängig von ihrer Eigenperiodendauer so lang andauert, daß der Phasenwert von 180° überschritten wird.

Der Ablösephasenwert einer Geste gibt an, bei welchem Phasenwert die gestische Aktivierung endet, d.h. welcher Realisierungsgrad von der zugehörigen Geste maximal erreicht wird. Je größer der Ablösephasenwert, um so geringer die realisierte relative Artikulator-Target-Differenz und damit um so höher der (maximale) Realisierungsgrad der entsprechenden Geste. Der Assoziationsphasenwert kennzeichnet die zeitliche Lage einer Geste relativ zu einer anderen. Hier gilt, daß mit ansteigendem Assoziationsphasen-

²Erläuterung zu Abb. 18.2. Durchgezogene und gestrichelte Linie: Zeitfunktionen gestisch gesteuerter Artikulationsbewegungen. Beide Gesten unterscheiden sich nur in der Eigenperiodendauer. Die gestrichelt dargestellte Zeitfunktion weist eine geringere Eigenperiodendauer auf (TE=120msec) als die der Geste mit der durchgezogenen dargestellten Zeitfunktion (TE=200msec). Abszisse: unten: absolute Zeitskala; oben: Phasenskalen für beide Gesten. Die Zuordnung dieser Skalen zu den Gesten erfolgt nach Linientyp (durchgezogene oder gestrichelte Linie). Ordinate: relative Artikulator-Target-Distanz. Die punktierte Linie deutet die Grenze zwischen Transitions- und Haltebereich einer Geste an.

wert der Grad der zeitlichen Überlappung der Geste relativ zu einer anderen zunimmt. Beispielsweise würde ein Ansteigen des Assoziationsphasenwertes der apikalen Beinahverschlußgeste in [mʊs] eine Verschiebung dieser Geste nach links gegenüber der vokalischen Geste der Äußerung bewirken (siehe Pfeil in Abbildung 18.1).

18.3 Die Realisierungen der Sprechtempoänderungen im gestischen Ansatz

18.3.1 Drei mögliche artikulatorisch-gestische Prozesse

Eine Erhöhung des Sprechtempos entspricht einer zeitlichen Kürzung der Gesamtdauer einer Äußerung. Diese Kürzung kann entweder durch Verringerung der zeitlichen Dauer einzelner (oder aller) Gesten der Äußerung oder durch Erhöhung der zeitlichen Überlappung der Gesten untereinander realisiert werden. Die Verringerung der zeitlichen Dauer einer Geste kann durch Verringerung der Eigenperiodendauer bei konstanter Ablösephase oder durch Verringerung der Ablösephase bei konstanter Eigenperiodendauer realisiert werden. Im ersten Fall bleibt die durch die Geste realisierte Annäherung des Artikulators an die Targetlage bei Verringerung der zeitlichen Ausdehnung der Geste erhalten. Im zweiten Fall wird bei Konstanz der Artikulatorgeschwindigkeit die Verringerung der zeitlichen Ausdehnung der Geste durch einen damit verbundenen geringeren (maximalen) Realisierungsgrad der Geste bezahlt. Der von der Geste gesteuerte Artikulator kann jetzt nur eine vergleichsweise geringere Annäherung an die Targetlage realisieren. Es entsteht artikulatorischer „Undershoot“ (Lindblom 1983, p. 228). Im Fall zweier auf den gleichen Artikulator wirkender Gesten kann die Verringerung der zeitlichen Ausdehnung einer Geste aufgrund der Verringerung der Ablösephase auch aus einer zunehmenden zeitlichen Überlappung dieser Geste mit einer zeitlich nachfolgenden Geste resultieren, so daß im folgenden nur zwischen dem Prozeß der Erhöhung der Artikulatorgeschwindigkeit (Verringerung der Eigenperiodendauer) und dem Prozeß der Erhöhung der zeitlichen Überlappung von Gesten (Verringerung der Ablösephase und/oder Erhöhung der Assoziationsphase) differenziert werden muß.

Bei ausreichend starker Erhöhung des Überlappungsgrades von Gesten können segmentale Änderungen resultieren (Browman und Goldstein 1990). Es können Vokalreduktionen zum Schwa, Assimilationen des Artikulationsortes und der Artikulationsart und auch Lautelisionen auftreten. Der Prozeß der Verringerung der Eigenperiodendauer allein führt hingegen nur zu einer Verkürzung der Lautdauern, ohne aber die artikulatorischen Bewegungsmuster qualitativ zu ändern.

18.3.2 Artikulatorische Messungen

Die Meßdaten wurden mittels der elektromagnetischen Artikulographie (Schönle 1988) erhoben. Diese Methode liefert die Ortskoordinaten von fünf Empfängerspulen in der Mediosagittalebene als Funktion der Zeit. Eine von uns entwickelte Methode zur Datenauswertung erlaubt die Abschätzung der gestischen Parameter (Kröger et al. 1995).

Das Sprachdatenmaterial umfaßt artikulatorische Daten von zwei Sprecherinnen (CO und DM). Die beiden Versuchspersonen sprachen die Äußerung „Ich habe [bV:bəbV:] betont. Ich sage [bV:bəbV:]“ mit den vier Langvokalen [V:] = [a:], [ɛ:], [e:], [i:]. Es wurde einmal die erste Silbe und einmal die letzte Silbe der dreisilbigen Logatome akzentuiert ([ˈbV:bəbV:] oder [bV:bəˈbV:]) und der Satz wurde einmal in normalem und einmal in schnellem Sprechtempo produziert. Aus jeder Äußerung wurden zwei vokalische Gesten, nämlich jeweils die der letzten Silbe des Logatoms, analysiert. Somit ergeben sich bei 6 Wiederholungen für vier Vokale, zwei Akzentuierungsmodi und zwei Sprechtempi insgesamt 96 Äußerungen mit 192 analysierten Gesten.

Für die Sprechtempoänderung wurden insbesondere die Eigenperiodendauer und der Phasenwert, der dem Endzeitpunkt des akustischen Vokalsegments entspricht, analysiert. Eine Verringerung dieses Endphasenwertes entspricht einer Erhöhung der zeitlichen Überlappung der Gesten. Die Messungen zeigten, daß die Änderungen sowohl der Eigenperiodendauer als auch der Endphase für beide Sprecherinnen signifikant sind. Es ergibt sich im Fall schneller gegenüber normaler Sprechgeschwindigkeit für beide Sprecherinnen eine Verringerung der Eigenperiodendauer und eine Verringerung des Endphasenwertes. Dabei verringert sich die Eigenperiodendauer relativ zum Gesamtmittelwert um 22% für Sprecherin CO und um 19% für Sprecherin DM, die Endphase um 8% für Sprecherin CO und um 7% für Sprecherin DM. Somit kann geschlossen werden, daß Sprechgeschwindigkeitserhöhung sowohl durch Verringerung der Eigenperiodendauer wie auch durch Änderung der gestischen Überlappung realisiert wird, wobei in diesem Korpus die Änderung der Eigenperiodendauer den stärksten Faktor der Sprechgeschwindigkeitsänderung darstellt. Der Prozeß der Änderung des zeitlichen Überlappungsgrades von Gesten tritt bei diesen artikulatorischen Messungen wahrscheinlich deshalb nicht stärker hervor, weil die Versuchspersonen sich bemühten, die oben angegebenen Logatome des Korpus ohne segmentale Reduktionen zu produzieren. Es ist aber festzuhalten, daß beide postulierten Artikulationsstrategien auftreten.

18.3.3 Perzeptive Untersuchungen

Anhand von perzeptiven Untersuchungen mittels unseres phonetischen Produktionsmodells (Kröger 1993) kann gezeigt werden, daß durch die Erhöhung des Überlappungsgrades von Gesten viele der diskreten segmen-

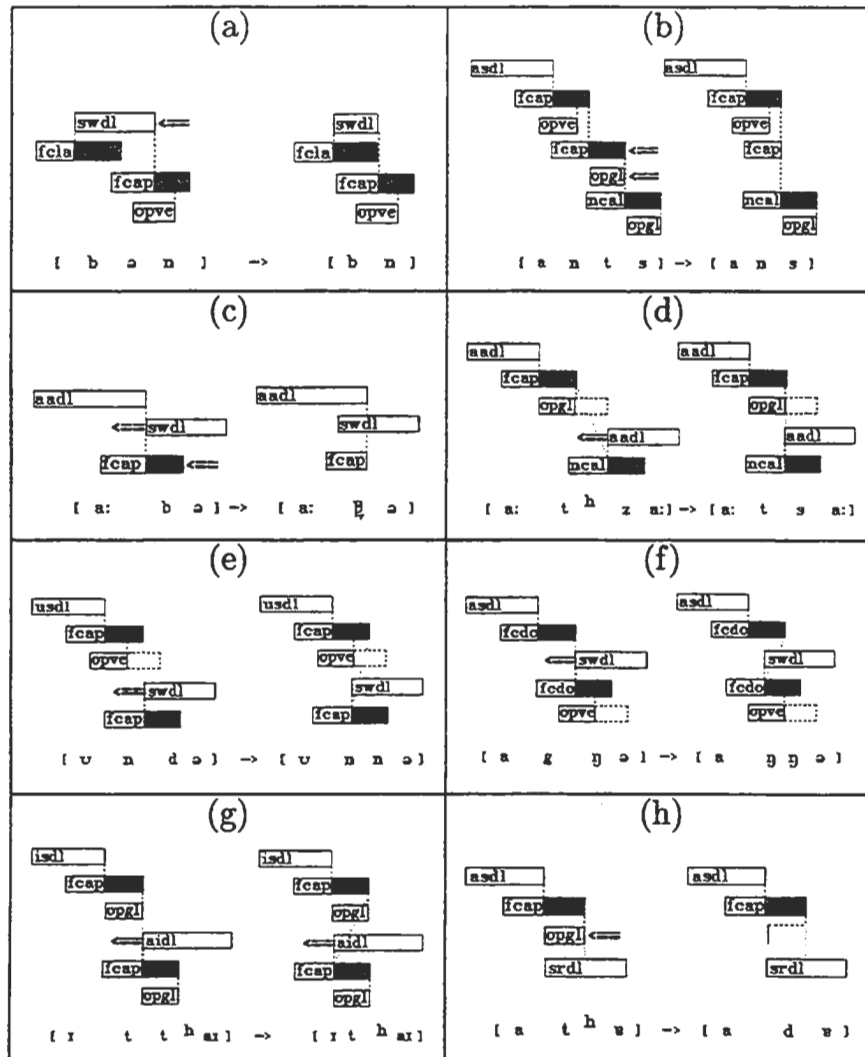


Abbildung 18.3: Gestische Partitur mit Kennzeichnung der gestischen Verschiebung.

talen Änderungen nach (Kohler 1995, S.205ff) entstehen. Hier wurde nur der Prozeß der Erhöhung der zeitlichen Überlappung von Gesten getestet, während die Eigenperiodendauern der Gesten nicht geändert wurden. Die Generierung einer diskreten segmentalen Änderung setzt voraus, daß mittels eines gestischen Verschiebungsprozesses in unserem Produktionsmodell zwei konkrete Formen der gestischen Partitur einer Äußerung generiert werden können, deren akustisches Korrelat die geforderte segmentale Änderung perzeptiv ergibt. Die qualitativen Perzeptionstests wurden in der Form durchgeführt, daß die akustischen Stimuli der beiden Formen der Äußerung von einem phonetisch geschulten Muttersprachler des Deutschen transkribiert wurden. Der Perzeptionstest wurde als erfolgreich angesehen, wenn die Transkription der beiden Formen der Äußerung die von Kohler (1995) vorgegebene segmentale Änderung enthielt.

Aufgrund einfacher gestischer Verschiebungen generierbare segmentale Änderungen

(a) Die *Schwa-Elision* (Kohler 1995, S.206f) wurde für die Beispiele „eben“ [bən] → [bn], „reden“ [dən] → [dn], „legen“ [gən] → [gn], und „Adel“ [dəl] → [dl] modelliert. Die zugehörige gestische Verschiebung ist für das Wort „eben“ dargestellt (Abbildung 18.3 (a)).³ In allen vier Fällen reicht die Verringerung der zeitlichen Ausdehnung der dorsal-labialen Schwaformenden Geste (swdl) der zweiten Silbe, d.h. die Verringerung des Wertes der Ablösephase dieser Geste auf RP=180° bis RP=100° aus, um die geforderten segmentalen Änderungen zu erzeugen. Diese Verringerung der zeitlichen Ausdehnung bewirkt eine Verringerung des zeitlichen Abstandes zwischen den Verschlußphasen der vor und hinter dem Schwa gelegenen konsonantischen Gesten (fcla und fcap in Abbildung 18.3 (a)), da die konsonantische Geste, die auf die Schwa-Geste folgt, zeitlich mit dem Ende der Schwa-Geste koordiniert ist. Die der Schwa-Geste nachfolgende konsonantische Geste, also im Beispiel „eben“ die apikale Vollverschlußgeste (fcap), und auch die hierzu assoziierten Gesten, also im Beispiel „eben“ die velische Öffnungsgeste (opve), werden bei Kürzung der Schwa-Geste hin zu früheren Zeitpunkten verschoben. Eine weitere Verringerung der zeitlichen Ausdehnung der Schwa-Geste bis zur vollständigen Reduktion dieser Geste (RP=0°) führt in allen angegebenen Beispielen zur völligen Überlappung der vor und hinter der Schwa-Geste auftretenden konsonantischen Gesten und damit zu sehr stark reduzierten Formen. So entsteht für das Wort „reden“ bei der weiteren Verringerung der zeitlichen Ausdehnung der Schwa-Geste von RP=100° auf RP=0° die segmentale Änderung [dn] → [n]. Ein vollständiger Wegfall der Schwa-Geste ist somit zur Elision des Schwa-Segmentes nicht erforderlich.

(b) Die */t/-Elision* (ebd., S. 208f) wurde für die Beispiele „Glanz“ [nts] → [ns], „erhältst“ [lts] → [ls], „restlich“ [stl] → [sl] und „rechtlich“

³Erläuterung zu Abb. 18.3. (a) Schwa-Elision in „eben“ (b) t-Elision in „Glanz“, (c) Änderung des Öffnungsgrades in „ich habe“, (d) progressiven Assimilation der Stimmlosigkeit in „ratsam“, (e) progressiven Assimilation der Nasalität t in „Bundes“, (f) regressiven Assimilation der Nasalität in „Agnes“, (g) Geminatenreduktion in „mitteilen“ und (h) Sonorisierung in „hat er“. Linke (rechte) Seite: gestische Partitur vor (nach) der gestischen Verschiebung. Die horizontale Ausdehnung der Balken repräsentiert die Länge des Aktivitätsintervalls der entsprechenden Geste. Die vertikalen Linien sind die Assoziationslinien. Im Fall der konsonantischen Gesten wurde das Halteintervall der Geste, d.h. der Zeitbereich der Ausbildung des Voll- oder Beinahverschlusses, schraffiert dargestellt; das Transitionsintervall enthält den Gestennamen. In den Abbildungen (d)–(f) markiert im Fall von Velums- oder glottalen Öffnungsgesten ein angehängter gestrichelter Balken die gesamte Zeitdauer einer Öffnung. Die gestischen Verschiebungen sind jeweils durch einen fettgedruckten Pfeil dargestellt. Steht dieser Pfeil hinter der Geste, so wird die zeitliche Ausdehnung der Geste verringert. Steht der Pfeil vor der Geste, so wird die Geste insgesamt verschoben.

[çtl] → [çl] modelliert. Die zugehörige gestische Verschiebung ist für das Wort „Glanz“ dargestellt (Abbildung 18.3 (b)). In allen vier Fällen reicht die Verringerung der zeitlichen Ausdehnung der apikalen Vollverschlußgeste (fcap) zur Produktion des [t], d.h. die Verringerung des Wertes der Ablösephase auf RP=180° bis RP=100°, zusammen mit der (fast) vollständigen zeitlichen Reduktion der zugehörigen glottalen Öffnungsgeste (opgl), d.h. der Verringerung des Wertes der Ablösephase dieser Geste auf RP=10° bis RP=0°, aus, um die geforderten segmentalen Änderungen zu erzeugen. Die Verringerung der zeitlichen Ausdehnung der Vollverschlußgeste des [t] bewirkt eine Verkürzung des zeitlichen Abstandes zwischen den Verschlußphasen des vorhergehenden und des nachfolgenden Konsonanten, da die konsonantischen Gesten innerhalb eines Konsonantenclusters miteinander zeitlich koordiniert sind. Ab einer Verringerung der zeitlichen Ausdehnung der Vollverschlußgeste des [t] unter RP=180° ist die Verschlußphase des [t] vollständig eliminiert, und es beginnt die zeitliche Überlappung der Verschlußphasen des vorhergehenden und des nachfolgenden Konsonanten – also im Beispiel „Glanz“ die Überlappung der Verschlußphasen des [n] mit der Verschlußphase des [s]. Eine weitere Reduktion der Vollverschlußgeste des [t] ist nicht nötig. Eine vollständige Reduktion dieser Geste würde eine vollständige Überlappung der vorhergehenden und der nachfolgenden konsonantischen Geste bewirken und zu im Deutschen nicht auftretenden reduzierten Formen führen. Die vollständige oder fast vollständige Reduktion der glottalen Öffnungsgeste des [t] ist erforderlich, da diese Geste mit der Kürzung der übergeordneten apikalen Vollverschlußgeste nach links verschoben wird. Ansonsten würde beispielsweise in „Glanz“ eine im Deutschen nicht auftretende Entstimmung des vorangehenden Nasals auftreten.

(c) Die *Änderung des Öffnungsgrades* (ebd., S. 209) wurde für die Beispiele „ich habe“ [b] → [β] und „ich lege“ [g] → [ɣ] modelliert. Die zugehörige gestische Verschiebung ist für den Satz „ich habe“ dargestellt (Abbildung 18.3 (c)). In beiden Fällen führt eine Verringerung der zeitlichen Ausdehnung der Vollverschlußgeste (fcla, Abbildung 18.3 (c)) auf RP=200° bis RP=180° zum perceptiven Eindruck des homorganen Approximanten. Bei einer noch stärkeren Reduktion der Vollverschlußgeste ist diese Geste perceptiv nicht mehr wirksam. Es entstehen dann reduzierte Formen, die im Deutschen nicht auftreten. Die Verschiebung der Schwa-Geste ist für die Beibehaltung der Dauer des Schwa-Lautes und für die zeitlich synchrone Realisierung des artikulatorischen Überganges vom [a:] zum Schwa zusammen mit der konsonantischen Konstriktionsgeste wichtig.

(d) Die *progressive Assimilation der Stimmlosigkeit* (ebd., S. 209) wurde für die Beispiele „ratsam“ [tz] → [ts], „dasselbe“ [sz] → [ss], „das Bad“ [sb] → [sβ] und „wegbringen“ [kb] → [kβ] modelliert. Die zugehörige gestische Verschiebung ist für das Wort „ratsam“ dargestellt (Abbil-

dung 18.3 (d)). Hier galt es, nicht die reduzierte Form, sondern die Ausgangsform zu erzeugen, da in unserem Produktionsmodell bereits eine enge zeitliche Verknüpfung benachbarter konsonantischer Gesten vorliegt. Dies führt im Fall des Wortes „ratsam“ (Abbildung 18.3 (d), rechte Hälfte) bereits zum Wegfall der Aspiration des [t] und zur Entstimmung des Frikativs. Die Realisierung der Ausgangsform (stimmloser Frikativ in „ratsam“) kann erreicht werden, indem der Grad der zeitlichen Überlappung zwischen den konsonantischen Gesten des [t] und des [z] verringert wird. Umgekehrt wird die progressive Assimilation der Stimmlosigkeit einfach durch eine Erhöhung des Grades der zeitlichen Überlappung dieser konsonantischen Gesten erreicht. Diese Änderung des Überlappungsgrades der konsonantischen Gesten resultiert aufgrund der zeitlichen Koordinierung der Gesten letztlich aus einer zunehmenden Überlappung der vokalischen Gesten (aadl, Abbildung 18.3 (d)).

(e) Die *progressive Assimilation der Nasalität* (ebd., S. 210) wurde für die Beispiele „umbenennen“ [mb] → [mm], „Bundes“ [nd] → [nn] und „angegeben“ [ŋg] → [ŋŋ] modelliert. Die zugehörige gestische Verschiebung ist für das Wort „(des) Bundes“ dargestellt (Abbildung 18.3 (e)). In allen drei Fällen ist eine zunehmende Überlappung der Verschlussgeste zur Realisierung des Nasals (fcap, Abbildung 18.3 (e)) und der Verschlussgeste des folgenden Plosivs (fcap, weiter unten in Abbildung 18.3 (e)) auf bis zu 50% der Dauer der Verschlussphase ausreichend, um die segmentale Änderung zu generieren. Diese zunehmende zeitliche Überlappung der konsonantischen Gesten kann auch im Fall der progressiven Assimilation der Nasalität ebenfalls allein auf die Verringerung des zeitlichen Abstandes der Silben und damit auf die zeitliche Verschiebung der Schwa-Geste der zweiten Silbe zurückgeführt werden (Abbildung 18.3 (e)). Eine weitergehende Überlappung der konsonantischen Gesten führt dann zur zeitlichen Kürzung der Geminaten ([nn] → [n]).

(f) Die *regressive Assimilation der Nasalität* (ebd., S. 210) wurde für die Beispiele „Agnes“ [gŋ] → [ŋŋ], „eben“ [bm] → [mm] und „werden“ [dn] → [nn] generiert. Für die Beispiele „eben“ und „werden“ ist die hier angegebene Ausgangsform bereits die nach erfolgter Schwa-Elision entstehende reduzierte Form. Die zugehörige gestische Verschiebung ist für das Wort „Agnes“ dargestellt (Abbildung 18.3 (f)). In allen vier Fällen entsteht die geforderte segmentale Änderung bei Durchführung der gleichen gestischen Verschiebungen wie im Fall der progressiven Nasalassimilation. Es ist eine zunehmende zeitliche Überlappung der konsonantischen Gesten (fcdo, Abbildung 18.3 (f)) bis auf 50% Überlappungsgrad für die Halteintervalle dieser Gesten durchzuführen.

(g) Die *Geminatenreduktion* (ebd., S. 210f) wurde für die Beispiele „kommen“ [mm] → [m], „mitteilen“ [tt] → [t] und „wegkommen“

[kk] → [k] generiert. Die zugehörige gestische Verschiebung ist für das Wort „mitteilen“ dargestellt (Abbildung 18.3 (g)). In allen drei Fällen entsteht eine vollständige Überlappung der die Doppelkonsonanten produzierenden konsonantischen Gesten (fcap, Abbildung 18.3 (g)). Im silbenorientierten gestischen Koordinierungsmodell resultiert die zugehörige gestische Verschiebung wiederum aus einer zunehmenden zeitlichen Annäherung der Silbenkerne und damit aus einer zunehmenden zeitlichen Überlappung der vokalischen Gesten (isdl, aidl, Abbildung 18.3 (g)).

(h) Die *Sonorisierung* (ebd., S. 211) wurde für die Beispiele „muß ich“ [s] → [z] und „hat er“ [t^h] → [d^h] modelliert. Die zugehörige gestische Verschiebung ist für den Satz „hat er“ dargestellt (Abbildung 18.3 (h)). In beiden Fällen ist eine (fast) vollständige Reduktion der glottalen Öffnungsgeste (opgl) ausreichend. Dabei wird allerdings von vornherein von einer reduzierten Formstufe ausgegangen, bei der die (in Abbildung 18.3 (h) nicht dargestellte) glottale Vollverschlußgeste zur Realisierung eines silbeninitialen [ʔ] der zweiten Silbe (des zweiten Wortes) bereits vollständig reduziert ist.

Aufgrund einfacher gestischer Verschiebungen nicht generierbare segmentale Änderungen

Schwa-Elision, /t/-Elision, Änderung des Öffnungsgrades, progressive Assimilation der Stimmlosigkeit, regressive und progressive Assimilation der Nasalität, Geminatenreduktion und Sonorisierung sind problemlos mittels der oben beschriebenen einfachen gestischen Verschiebungen generierbar. Allerdings war es nicht möglich, alle segmentalen Änderungen der progressiven und der regressiven Assimilation des Artikulationsortes und der regressiven Assimilation der Artikulationsart auf diese Weise zu erzeugen. Es wird angenommen, daß die Produktion dieser segmentalen Änderungen nicht durch gestische Verschiebungen, sondern durch „artikulatorische Reorganisation“ (Gay 1981) verursacht wird.

(a) Die *regressive Assimilation der Artikulationsart* (ebd., S. 209) wurde für die Beispiele „das Schiff“ [sf] → [ʃf] und „Eissschrank“ [sf] → [ʃf] modelliert. Die in beiden Beispielen geforderte segmentale Änderung ist realisierbar, indem die alveolare Beinahverschlußgeste zur Realisierung des [s] aus „das“ vollständig reduziert wird. Allerdings ist dies kein Assimilationsprozeß unter Bildung des oben angegebenen Doppelkonsonanten. Hierzu wäre die weiter unten beschriebene artikulatorische Reorganisation in der Form des Austausches der alveolaren gegen eine postalveolare Beinahverschlußgeste nötig.

(b) Die *progressive Assimilation des Artikulationsortes* (ebd., S. 208) wurde im Fall der Assimilation von Nasalen für die Beispiele „eben“ [bn] → [bm],

„verlogen“ [gn] → [gŋ], „trocken“ [kn] → [kŋ], „kommen“ [mn] → [mm], „fangen“ [gn] → [gŋ] und im Fall der Assimilation von Plosiven für die Beispiele „Hemden“ [mdn] → [mbm], „Beamten“ [mtn] → [mpm] und „Gegenden“ [gndn] → [gŋgŋ] modelliert. Als einfacher gestischer Verschiebungsprozeß kommt hier nur die Reduktion der Schwa-Geste und damit die Zunahme der zeitlichen Überlappung der der Schwa-Geste folgenden konsonantischen Geste mit der der Schwa-Geste vorhergehenden konsonantischen Geste in Betracht. Hier ergibt sich aber keine gleichartige Transkription für alle Beispiele: Während in den genannten Beispielen die Schwa-Elision eindeutig perzipiert wird, wird abhängig vom Beispiel teils die Ortsassimilation perzipiert – z.B. bei „eben“ und „trocken“ – und teils nicht – z.B. bei „verlogen“ oder „kommen“.

(c) Die *regressive Assimilation des Artikulationsortes* (ebd., S. 207) wurde für die Beispiele „anbinden“ [nb] → [mb], „geht baden“ [tb] → [pb], „angeben“ [ng] → [ŋg] und „mitkommen“ [tk] → [kk] untersucht. Als einfacher gestischer Verschiebungsprozeß kommt hier nur die Zunahme der zeitlichen Überlappung zweier Silben und daraus resultierend die zunehmende zeitliche Überlappung der entsprechenden, im Auslaut der ersten Silbe und im Anlaut der zweiten Silbe stehenden konsonantischen Gesten in Betracht. Hier wird in keinem der modellierten Beispiele der Assimilationseffekt perzipiert. Es wurde für jedes Beispiel der in der Ausgangsform auftretende apikale Artikulationsort auch in der Transkription der reduzierten Form festgehalten.

Darüber hinaus wurden die Beispiele „mit meinem“ [nm] → [mm], „mit jedem“ [dm] → [bm], und „mit fettem“ [tm] → [pm] (Kohler 1990, S.85) untersucht. Auch hier wurde die zu erwartende regressive Ortsassimilation nicht festgestellt. Dafür entsteht bei einigen Formen der Eindruck einer progressiven Assimilation des Artikulationsortes („mit meinem“ [nm] → [nn], „mit jedem“ [dm] → [dn], „mit fettem“ [tm] → [tn]). Eine derartige progressive Assimilation des Artikulationsortes zugunsten apikaler Laute tritt aber nach den von Kohler (1990, 1995) gegebenen Regeln im Deutschen nicht auf.

Quantitative perzeptive Analyse zur regressiven und progressiven Assimilation des Artikulationsortes

Zur genaueren Untersuchung der progressiven und der regressiven Ortsassimilation wurden die Beispiele „mit meinem“, „mit jedem“ und „mit fettem“ einer quantitativen perzeptiven Analyse unterzogen. In Voruntersuchungen wurde festgestellt, daß die regressive Ortsassimilation der Labiallaute nur dann eindeutig perzipiert wird, wenn die apikale Vollverschlußgeste des dem Schwa-Laut vorhergehenden Konsonanten durch eine entspre-

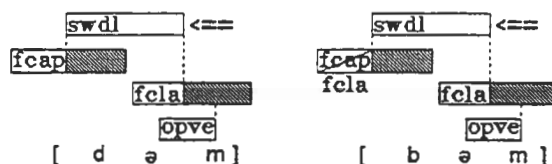


Abbildung 18.4: Gestische Partituren der beiden Ausgangsformen für die letzte Silbe von „mit jedem“.

chende labiale Geste ersetzt wird (Abbildung 18.4).⁴ Dies ist ein gestischer Reorganisationsprozeß.

Für die quantitativen perzeptiven Analysen wurden zu jedem der drei Beispiele jeweils eine nicht reorganisierte und eine reorganisierte Ausgangsform generiert, und dann beide derselben gestischen Verschiebung unterzogen: Verringerung der zeitlichen Ausdehnung der dorsal-labialen Schwa-Geste (swdl). Dies führt zu einer zunehmenden zeitlichen Überlappung der konsonantischen Gesten dieser Silbe. Es wurden zu jeder Form sieben Stimuli synthetisiert (Ablösephasenwert: RP=300⁰, 250⁰, 200⁰, ..., 0⁰), wobei die erste Form (RP=300⁰) eine gerade beginnende und die siebte Form (RP=0⁰) eine vollständige zeitliche Überlappung der Aktivitätsintervalle der konsonantischen Gesten aufweist.

Zwei Tests wurden mit 12 Hörern (Erst- bis Drittsemester des Studienschlusses Phonetik; Muttersprachler des Deutschen) durchgeführt, bei denen jeweils unterschiedliche Fragestellungen zu beantworten waren. Beim ersten Test sollte entschieden werden, ob der Konsonant vor dem Schwa in „mit meinem“ als /m/ oder /n/, in „mit jedem“ als /d/ oder /t/ und in „mit fettem“ als /t/ oder /d/ gehört wird. Beim zweiten Test sollte entschieden werden, ob der Konsonant nach dem Schwa als /m/ oder als /n/ wahrgenommen wird. Tabelle 18.1 gibt den Prozentsatz der Entscheidungen für den Artikulationsort „labial“ für beide Ausgangsformen (reorganisiert und nicht reorganisiert) ohne Überlappung und bei vollständiger Überlappung der konsonantischen Gesten wieder.

Die Ergebnisse zeigen, daß auch im Fall der reorganisierten Formen nicht zu annähernd 100% der labiale Artikulationsort perzipiert wird. Dies liegt wahrscheinlich an der Tatsache, daß die nicht-reduzierten reorganisierten Formen [mitmaeməm], [mitje:bəm], [mitfɛpəm] im Deutschen nicht auftreten und damit vom Hörer auch nicht erwartet werden.

Im Fall der nicht reorganisierten Form treten zwar signifikante Übergänge von einer phonemischen Kategorie zur anderen auf; allerdings sind sie in den hier durchgeführten Tests 1 und 2 (mit Ausnahme von Test 2

⁴Erläuterung zu Abb. 18.4. Links: nicht reorganisierte Form, rechts: reorganisierte Form und Reorganisationsprozeß. Der Pfeil kennzeichnet die für beide Formen durchgeführte gestische Verschiebung. Die Verschlussintervalle der konsonantischen Gesten wurden schraffiert.

Tabelle 18.1: Klassifikationsergebnisse in Prozent für „labial“ mit/ohne Überlappung in beiden Tests für nicht reorganisierte (nr) und für reorganisierte (ro) Formen. (Mittelwertbildung über jeweils zwei Formen: RP=0⁰ und RP=50⁰ bei voller Überlappung; RP=250⁰ und RP=300⁰ bei fehlender Überlappung).

	Keine Überlappung				Vollständige Überlappung			
	Test 1		Test 2		Test 1		Test 2	
	nr	ro	nr	ro	nr	ro	nr	ro
meinen	3	67	89	94	70	71	35	92
jeden	0	77	96	96	49	64	10	64
fetten	3	91	93	99	53	86	40	82

für „mit jedem“) eher tendenziell, nicht aber vollständig. So ist in beiden Tests trotz zunehmender Überlappung zwar eine Tendenz in Richtung perzeptiv wahrgenommener Ortsassimilation festzustellen, aber auch bei vollständiger Überlappung der Gesten liegt zumeist kein vollständiger perzeptiver Übergang zur entsprechend anderen phonemischen Kategorie vor. Im Einklang mit den Ergebnissen der qualitativen Perzeptionstests zeigt Test 2 bei den nicht reorganisierten Formen sogar eine im Deutschen nicht auftretende (progressive) Assimilation des Artikulationsortes zugunsten von apikalen Konsonanten. Es werden hier mit hohen Prozentsätzen die Formen [mitmaenn], [mitje:nn] und [mitfenn] perzipiert. Bei den reorganisierten Formen (insbesondere bei „mit meinem“ und „mit fettem“) bleibt aber der labiale Artikulationsort unabhängig vom Überlappungsgrad erhalten. Im Fall des Überganges von nicht-reorganisierten zu reorganisierten Formen entsteht somit die Assimilation zugunsten des labialen Artikulationsortes. Dieser Test weist damit auf die Notwendigkeit gestischer Reorganisation zur Modellierung der Assimilation des Artikulationsortes hin.

18.4 Schlußbemerkung

Es wurde anhand von artikulatorischen Messungen gezeigt, daß Sprechtempoerhöhung aufgrund unterschiedlicher Strategien (Verringerung der gestischen Eigenperiodendauer oder Erhöhung der zeitlichen Überlappung von Gesten) realisiert werden kann. Anhand von perzeptiven Untersuchungen wurde gezeigt, daß die mit Sprechtempoerhöhung verbundenen diskreten segmentalen Änderungen in vielen Fällen (Schwa-Elision, /t/-Elision, Änderung des Öffnungsgrades, progressive Assimilation der Stimmlosigkeit, progressive und regressive Assimilation der Nasalität, Geminatenreduktion und Sonorisierung) aufgrund der Zunahme gestischer Überlappung ohne Wegfall und ohne weitergehende qualitative Änderungen einzelner Gesten entstehen. In einigen Fällen (regressive Assimilation der Artikulationsart, progressive und regressive Assimilation des Artikulationsortes) tritt aller-

dings artikulatorische Reorganisation in Form von Austausch des die Geste realisierenden Artikulators auf. Auch diese Reorganisation kann von unserem Produktionsmodell realisiert werden.

Literaturverzeichnis

- C. Browman und L. Goldstein (1986). Towards an articulatory phonology. *Phonology Yearbook* 3: 219–252.
- C. Browman und L. Goldstein (1989). Articulatory gestures as phonological units. *Phonology* 6: 201–251.
- C. Browman und L. Goldstein (1990). Tiers in articulatory phonology, with some implications for casual speech. In: M. B. J. Kingston, Hg., *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, S. 341–376. Cambridge University Press, Cambridge. Auch in: Haskins Laboratories Status Report on Speech Research SR-92 (1987), 1–30.
- C. Browman und L. Goldstein (1992). Articulatory phonology: An overview. *Phonetica* 49: 155–180.
- O. Fujimura (1986). Relative invariance of articulatory movements: An iceberg model. In: D. K. J.S. Perkell, Hg., *Invariance and Variability in Speech Processes*, S. 226–242. Lawrence Erlbaum, Hillsdale, New Jersey.
- T. Gay (1981). Mechanisms in the control of speech rate. *Phonetica* 38: 148–158.
- J. Kelso, E. Saltzman und B. Tuller (1986). The dynamical perspective on speech production: Data and theory. *Journal of Phonetics* 14: 29–59.
- K. Kohler (1990). Segmental reduction in connected speech in German: Phonological facts and phonetic explanations. In: A. M. W.J. Hardcastle, Hg., *Speech Production and Speech Modelling*, S. 69–92. Kluwer Academic Press, Dordrecht.
- K. Kohler (1995). *Einführung in die Phonetik des Deutschen*. 2. Auflage. Erich Schmidt Verlag, Berlin.
- B. Kröger (1993). A gestural production model and its application to reduction in German. *Phonetica* 50.
- B. Kröger, G. Schröder und C. Opgen-Rhein (1995). A gesture-based dynamic model describing articulatory movement data. *Journal of the Acoustical Society of America* 98: 1878–1889.
- A. Liberman und I. Mattingly (1985). The motor theory of speech perception revised. *Cognition* 21: 1–36.
- B. Lindblom (1983). Economy of speech gestures. In: P. MacNeilage, Hg., *The Production of Speech*, S. 217–246. Springer-Verlag, New York.
- E. Saltzman (1985). Task dynamic coordination of the speech articulators: A preliminary model. Status Report on Speech Research SR-84, Haskins Laboratories. pp. 1–18.
- E. Saltzman und K. Munhall (1989). A dynamic approach to gestural patterning in speech production. *Ecological Psychology* 1: 333–382.
- P. Schönle (1988). *Elektromagnetische Artikulographie*. Springer-Verlag, Berlin.
- M. Walther und B. J. Kröger (1994). Phonologie-Phonetik-Kopplung in einem constraintbasierten gesturalen Modell. In: H. Trost, Hg., *Tagungsband Konvens '94. Verarbeitung natürlicher Sprache*.