

GERHART LINDNER

Sektion Rehabilitationspädagogik und
Kommunikationswissenschaft, Humboldt-Universität
Berlin (DDR) 1040

ZUSAMMENFASSUNG

Die Entwicklung der Computertechnik macht es möglich, die artikulatorischen Bewegungen beim Sprechen auf einem Monitor zu modellieren. Damit sie in der Lehre im Hochschulunterricht eingesetzt werden können, ist ein Programm, das sich mit Hilfe eines Kleincomputers darbieten läßt, von Vorteil. Zu einem Grundmuster des Kopfes werden die Bewegungen der Artikulationsorgane (Lippen, Unterkiefer, Gaumensegel) mit differenzierten Programmen ausgewiesen. Bei Eingabe der Lautfolge werden sie zusammengefügt. Die Zungenlinie wird durch einen besonderen Algorithmus ermittelt, in dem die Laute fixierte Höhepunkte darstellen. Um den Bewegungsablauf auch in den Zwischenphasen wirklichkeitsgetreu abbilden zu können, ist der Vergleich mit Real-aufnahmen notwendig. Für die deutsche Sprache wird das Ergebnis der Modellierung mit Röntgen-Zeitlupen-Aufnahmen verglichen, die bereits ausgewertet sind und an denen einige Gesetzmäßigkeiten des koartikulatorischen Bewegungsablaufs ermittelt wurden. Die Optimierung des Programms für die Zungenlinie erfolgt in drei Schritten: 1. Lineare Interpolation, 2. Berücksichtigung des differenzierten Bewegungstempos der Artikulationsorgane, 3. Berücksichtigung der lautübergreifenden Koartikulation.

1. NOTWENDIGKEIT DER MODELLIERUNG ARTIKULATORISCHER BEWEGUNGEN

Das Sprechen ist ein weitgehend automatisierter Prozeß, bei dem eine Vielzahl von dezentralisiert im Organismus gelegenen Organen harmonisch zusammenwirkt. Bei der Anbildung oder der Korrektur des Sprechens ist es notwendig, in diesen automatisierten Prozeß einzugreifen. Um rationell vorgehen zu können, ist es erforderlich, diejenigen Artikulationsbewegungen zu erhalten und zu nutzen, die richtig ausgeführt werden, und gleichzeitig diejenigen Bewegungen zu korrigieren, die fehlerhaft sind. Dazu muß der Pädagoge über differenzierte Vorstellungen und detaillierte Einsichten in den Prozeß der miteinander verflochtenen Bewegungen der Sprechorgane verfügen. MENZERATH hat dieses komplexe Geschehen anschaulich als Sprechbewegungsgefüge bezeichnet./1/

Den Sprechbewegungsablauf darzustellen und zu lehren ist deshalb besonders schwierig, weil

- nur Höhepunkte des als Gesamtablauf über das akustische Klangprodukt kontrollierten Komplexes, die Laute, bewußt werden,
- die Umsetzung der Lautsprache in die Schrift nur diese Höhepunkte nutzt und damit die Orientierung des Kenntniserwerbs in bezug auf Ausschnitte aus dem Gesamtkomplex unterstützt,
- sich die Organe nicht, den isolierten Lauten entsprechend, plötzlich und ruckartig bewegen, sondern im Verlauf eines harmonischen Bewegungsgeschehens die Positionen durchlaufen, die den Lauten entsprechen,
- sich gerade zwischen den als Lauten gekennzeichneten Höhepunkten wichtige Bewegungen einzelner Organe vollziehen,
- das Tempo der Bewegungen der einzelnen Organe unterschiedlich ist, dies aber nur in den akustisch wirksamen Gesamtprozeß eingeht,
- sich die innere Anschauung über das Bewegungsgefüge nicht aus der Selbstbeobachtung gewinnen läßt, da wesentliche Bewegungen der unmittelbaren Beobachtung entzogen sind.

Zum Zweck der Korrektur und der systematischen Anbildung muß der Pädagoge in der Lage sein, den komplexen Bewegungsablauf in seine Einzelheiten aufzulösen. Die Vermittlung dieser Vorstellungen ist ebenfalls schwierig, weil

- die visuelle Beobachtung der Sprechorgane auf die Bewegungen von Lippen, Unterkiefer und Zungenspitze beschränkt ist,
- sich nur wenige Organe taktil kontrollieren lassen,
- die auditive zeitliche Differenzierung wohl zur Erkennung der Laute, nicht aber der Lautübergänge ausreicht,
- sich die Bewegungen, die sich in stimmlosen Perioden des Sprechens vollziehen, weder auditiv noch meßtechnisch über das akustische Signal erfassen lassen. Mit den phonetischen Anschauungsmitteln, wie sie sich in Lehrbüchern finden, lassen sich zwar Kenntnisse über die lautbezogenen Organpositionen vermitteln./2,3,4/ Diese Kenntnisse betreffen aber nur Höhe-

punkte des miteinander verflochtenen Bewegungsablaufs und nehmen auf die Veränderungen bei der Koartikulation keinen Bezug. Für die effektive Korrektur des zusammenhängenden und die Anbildung des fließenden Sprechens ist es aber notwendig, daß der Pädagoge weiß, wie sich die Bewegungen der Sprechorgane zwischen den Ausschnitten vollziehen, damit er beim Schüler die notwendigen Bewegungen stimulieren, entwickeln und kontrollieren kann. Denn der Schüler muß beim Sprechen Bewegungen vollziehen, und der Lehrer muß diese bewerten.

Für die Lehre ist es deshalb notwendig, ein anschauliches Modell zu entwickeln, mit dem sich die Bewegungen der Artikulationsorgane beim Sprechen demonstrieren lassen. Dadurch wird der Lernprozeß, der zu anwendungsfähigen Vorstellungen führt, abgekürzt und gleichzeitig über das Niveau hinausgeführt, das mit den heutigen Anschauungsmitteln erreichbar ist.

2. VORAUSSETZUNGEN FÜR DEN LÖSUNGSANSATZ

Damit die Aufgabe, die mit Hilfe der Computergrafik lösbar geworden ist, realisiert werden kann, müssen eine Reihe von Vereinfachungen vorgenommen werden. Sie betreffen

- den Verzicht auf die Individualität,
- die Darstellung der Bewegungen in nur zweidimensionaler Abbildung,
- die Beschränkung auf die deutsche Standardaussprache, wobei spätere Erweiterungen auf Sprechfehler, Dialekte oder Fremdsprachen vorgesehen werden,
- die Kontinuität der zeitlichen Auflösung, da die Bewegungen ohnehin auf einem Monitor dargestellt werden.

Damit eine computergrafische Modellierung vollzogen werden kann, werden die Einzelheiten des Kopflängsschnittes nach bewährtem Muster in unbewegliche und bewegliche Organe unterteilt./5/ Die unbeweglichen Teile dienen zum genauen Verfolgen der Bewegungen und werden unverändert beibehalten. Sie sind der konstante Teil des grafischen Programms:

- Oberkiefer mit Zähnen, Ansatz zum Nasenraum, hintere Rachenwand, Kehlkopf mit Stimmlippen in einer Mittelstellung. Die artikulatorischen Bewegungen während des Sprechens werden von folgenden Organen dargestellt:
- Unterkiefer mit Zähnen und Kinnlinie, Oberlippe, Unterlippe, Gaumensegel und Zungenlinie.

Die Einstellungen und Bewegungen der Stimmlippen sind in der zweidimensionalen Darstellung des Kopflängsschnittes nicht darstellbar. Für eine Weiterentwicklung ist geplant, sie in einer anderen Abbildungsebene (Draufsicht) in einer Ecke des Bildes einzublenden. Die aktuelle Lautfolge der modellierten Bewegungen wird in phonetischer Umschrift dargestellt.

3. PRINZIPIELLE SCHWIERIGKEITEN EINER ARTIKULATORISCHEN BEWEGUNGSSYNTHESE

Die Modellierung artikulatorischer Abläufe ist deshalb schwierig, weil es sich um die Darstellung von Einzelheiten handelt, die von der Phonetik bisher nicht oder kaum erarbeitet worden sind, da sie praktisch nicht gebraucht wurden. Sowohl bei der Anbildung und Korrektur des Sprechens als auch im Fremdsprachenunterricht können sich Lehrer und Schüler auf voll eingeübte Bewegungsautomatismen stützen. Auch bei einem völlig Stummen sind Bewegungsvollzüge der Organe, die zum Sprechen notwendig sind, für den rein emotionalen Ausdruck und zur Nahrungsaufnahme in einer bestimmten Weise eingespielt. Diese Bewegungen müssen im Verlauf des Lernprozesses umgestellt und anders koordiniert werden. Bei der Ausführung dieser natürlichen Bewegungen der Sprechorgane sind die einzelnen Teile, die im Modell separat dargestellt werden müssen, zwangsläufig miteinander verbunden und in ihrem Bewegungsinventar aufeinander abgestimmt.

Für die Modellierung der Bewegungen fallen diese Bedingungen weg, die das Bewegungsinventar einschränken. Außerdem entfällt die Möglichkeit, das auf dem Monitor Modellierte durch den akustischen Effekt zu kontrollieren. Für die Modellierung gibt es bisher keine Vorgaben für Grenzbedingungen der Einstellungen oder Bewegungen einzelner Organe und für deren Zusammenwirken; denn solche Angaben werden für den Unterrichtsprozeß nicht gebraucht. Deshalb steht die Modellierung von artikulatorischen Bewegungen heute vor Schwierigkeiten, die denen gleichen, die die akustische Sprachsynthese zu Beginn ihrer Arbeiten überwinden mußte. Das System für die Sprechbewegungssynthese ist ein offenes System. Mit ihm sind prinzipiell alle bildlichen Darstellungen der Sprechorgane realisierbar, auch solche, die auf den ersten Blick als unsinnig erkannt werden (z. B. wenn die Zunge aus zwei Teilen besteht). Die Schwierigkeit besteht darin, daß die Darstellung exakt der Wirklichkeit entspricht, damit dem Schüler wirklichkeitsgetreue und anwendbare Vorstellungen von Bewegungsvollzügen vermittelt werden.

4. VERFAHREN DES LÖSUNGSWEGES

Damit ein Gesamtbild modelliert werden kann, muß es nach bewährten Grundsätzen, die auch für bewegte Trickzeichnungen angewendet werden, in einzelne Teile zerlegt werden. Als Teilbilder werden benutzt:

- Das Grundmuster des Kopflängsschnittes. Es wurde nach anatomischer Vorgabe entworfen./6/
- Die verschiedenen Einstellungen des Unterkiefers. Zwischen maximaler Weite und minimaler Enge wurden weitere 7 Stellungen vorgesehen. Als 10. Position ist eine retrahierte Stellung geplant, um die labialen Engelaute darzustellen.

- Die aktiven Einstellungen von Ober- und Unterlippe. Dabei setzt die Oberlippe an feststehenden Punkten des Grundmusters an. Die Bewegungen der Unterlippe sind von der Position des Unterkiefers abhängig. Obwohl bekannt ist, daß in bestimmten artikulatorischen Abläufen die aktiven Bewegungen von Ober- und Unterlippe nicht übereinstimmen, wurden sie auf der Grundlage einer gemeinsamen Aktivität modelliert.

- Die verschiedenen Stellungen des Gaumensegels, dessen Darstellung an Festpunkten des Grundmusters ansetzt.

- Für die Erweiterung des gegenwärtigen Programms ist auch die Modellierung der Glottiseinstellungen vorgesehen. Dabei ist zu berücksichtigen, daß neben der Stimmstellung auch die Einatmungs-, Hauch-, Flüster- und Ruhestellung zu sehen sind. Die Bewegungen von Unterkiefer, Ober- und Unterlippe, Gaumensegel und auch der Glottis lassen sich mit Hilfe eines relativ kleinen Programms modellieren, da sich diese Bewegungen immer in der gleichen Weise wiederholen und die Variationsmöglichkeiten gering sind.

Größere Schwierigkeiten bereitet die Modellierung der Zungenlinie, da sie einerseits von den Bewegungen des Unterkiefers abhängig ist und andererseits in mindestens zwei Hauptabschnitte (Zungenspitze und -rücken) aufgeteilt werden muß. Um diese komplizierte Bewegung mit vielen Übergangsmöglichkeiten modellieren zu können, wurde ein spezielles Koordinatensystem entwickelt.

Es hat sich als ausreichend erwiesen, die Zungenlinie mit 15 Punkten festzulegen und die Zwischenräume durch gerade Strecken miteinander zu verbinden. Diese Punkte liegen an den Stellen maximaler Krümmung. Da bei den Vokalen der gesamte Hohlraum des Ansatzrohres an der Klangbildung beteiligt ist wurde entschieden, alle Werte der Zungenlinie fest zu speichern. Bei den Konsonanten wurde eine Differenzierung vorgenommen, indem jene Gebiete der Zungenlinie, die unbedingt eingenommen werden müssen, von solchen unterschieden werden, die bei der Lautbildung koartikulatorisch veränderlich sind.

Das Programm wurde von Anfang an so konzipiert, daß es für weitere Laute ergänzungsfähig ist und daß es in einem späteren Stadium mit dem Sprachsynthesator ROSY 4201 /7/ zusammenschaltet werden kann, so daß dann die Möglichkeit gegeben ist, die visuelle Synthese mit der akustischen zu kombinieren.

Aus den Teilbewegungen von Unterkiefer, Lippen, Gaumensegel und Zungenlinie läßt sich die beim Sprechen einer beliebigen Lautfolge notwendige Bewegung der Artikulationsorgane als Ganzes modellieren. Dabei reicht es aus, die Bewegungen, die sich in dem Zeitraum von 20 ms vollziehen, zu einem Bild zusammenzufassen. Eine feinere Unterteilung ist weder nötig noch

möglich, da auf dem Monitor pro Sekunde nur 50 Halbbilder gezeigt werden. Für den Zeittakt wurden die folgenden Vorgaben konzipiert, um die Wirklichkeitstreue der Modellierung zu überprüfen. Die Vorgaben für den Zeittakt wurden aus der Literatur entnommen /8/ und grob vereinfacht. Die Zeittakteinheit (ZTE) wurde als 20 ms festgelegt. Es beträgt die Dauer

- eines kurzen Vokals	4 ZTE
- eines langen Vokals	7 ZTE
- eines Diphthongs	7 ZTE
- des Murrelvokals	3 ZTE
- eines Konsonanten	5 ZTE

Diese Zeitvorgaben werden in Abhängigkeit von der Akzentuierung modifiziert:

- Zuschlag zu einer betonten Silbe 3 ZTE
- Kürzung bei einer unbetonten Silbe 2 ZTE

Der Vergleich mit dem real produzierten Sprachmaterial wurde an dem gleichen Testsatz vollzogen, der röntgenkinematografisch aufgenommen und ausgewertet worden war. /9/ Dieser Testsatz, der die häufigsten deutschen Lautfolgen enthält, lautet: "Wie denn, ist das einer der Steine, die ich im Winter anderwärts gefunden habe?" Die Zeit für die Modellierung ergibt nach den oben angegebenen Zeittaktwerten eine Dauer von 5,36 s und liegt damit innerhalb der individuellen Variationsbreite.

5. STANDORTBESTIMMUNG FÜR DIE REALISIERUNG

Wenn man davon ausgeht, daß die Laute Höhepunkte des artikulatorischen Geschehens sind, dann ist es möglich, die für die Laute entwickelten Gesamtbilder als die Ausgangspunkte für die Modellierung der Übergänge zu nutzen. Es wird zunächst davon abgesehen, daß auch die Lautpositionen im zusammenhängenden Sprechen veränderlich sind.

Es ist bekannt, daß vor allem die Konsonanten, aber auch die Vokale, koartikulatorischen und akzentabhängigen Veränderungen unterliegen, was sich besonders in der Realisierung unbetonter Silben der Rede ausdrückt. Sie sind für das Russische besser aufgearbeitet als für das Deutsche. /10/

Für das Deutsche liegen sie weder in algorithmisch nutzbarer Form noch in verwertbaren Abbildungsunterlagen vor. Deshalb wird bei der Modellierung zunächst an unveränderlichen Lautbildern festgehalten. Sie stellen im Modell die Phasen dar, die die Artikulationsorgane bei der Aussprache einnehmen oder wenigstens durchlaufen. Bei den langen Anteilen eines Lautes wird dann die für den Laut vorgegebene Position eine Anzahl von Phasen gleichgehalten. Bei den kurzen Lauten wird diese Phase zwar als Zielposition für die Modellierung zum Höhepunkt hin verwendet, aber nach deren Erreichen sofort als Ausgangsposition für die Ansteuerung zum nächsten Laut verwendet. Der Betrachter erlebt dann diese Position eines kurzen Lautes nur

als Durchgangspphase. Als einfachste Möglichkeit, die Übergänge zwischen den Lauten zu realisieren, bietet sich die lineare Interpolation an. Dieses Verfahren ist rechnerisch unkompliziert. Damit die Veränderungen, die sich innerhalb einer Lautfolge vollziehen, gut perzipierbar sind, werden, ehe sich das Bild zu verändern beginnt, die auszuführenden Veränderungen als Vektoren abgebildet. (Abb.1)

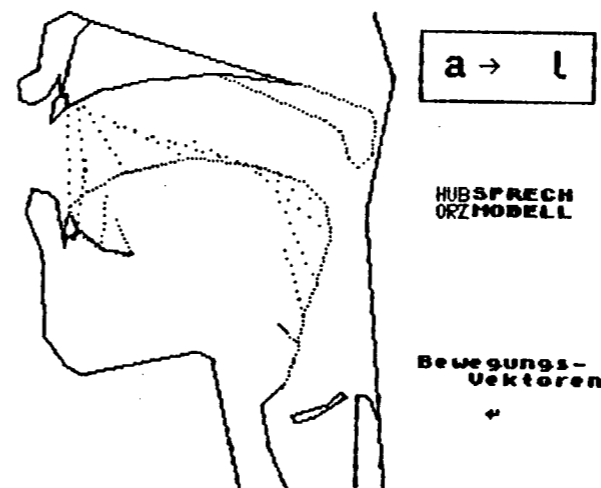


Abb. 1: Computerausdruck eines Positionsbildes mit den Bewegungsvektoren zum Folgelaut.

Die zweite Möglichkeit ist, die unterschiedliche Bewegungsgeschwindigkeit der Organe zu berücksichtigen. Zungenspitze und Lippen sind schnell-, Unterkiefer und Gaumensegel sind langsambewegliche Organe. Mit diesem Algorithmus, der vor allem die hohe Beweglichkeit der Zungenspitze berücksichtigt, wurde zeichnerisch die Lautfolge "der Steine" modelliert. Dies ist ein Ausschnitt aus dem Testsatz und liegt als röntgenkinematografische Analyse von verschiedenen Versuchspersonen vor. Der Vergleich zwischen Modellierung und Realisation erfolgte mittels des Motogramms. /11/ Der Vergleich zeigt gute Übereinstimmung zwischen Modellierung und Realisation. Die Modellierung liegt innerhalb der individuellen Variationsbreite. Die dritte Stufe der Annäherung an die Wirklichkeit ist noch perspektivisch. Sie besteht in der Berücksichtigung der koartikulatorischen Bedingungen, wobei berücksichtigt werden muß, daß die Konsonanten den stärkeren koartikulatorischen Veränderungen ausgesetzt sind.

6. MIT DEM PROGRAMM ERZIELTE ERGEBNISSE

Das Programm wurde mit dem Ziel ausgearbeitet, daß es als Lehr- und Unterrichtsmittel verwendet werden kann. Deshalb wurde es für den an allen Universitäten und Hochschulen der DDR vorhandenen Kleincomputer KC 85/2 ausgearbeitet. Damit ist be-

wiesen, daß ein so umfangreiches und kompliziertes Programm, wie es für die Bewegungssynthese notwendig wird, mit einem Kleincomputer realisierbar ist. Für einen Teil der Studenten wurde dadurch, daß das Programm in den planmäßigen Lehrprozeß des Phonetikunterrichts einbezogen wurde, die Einsicht in die Dynamik der Physiologie des Sprechens erheblich verbessert, die eigene Beschäftigung mit artikulatorischen Bewegungsvorgängen wesentlich stimuliert.

Die Mängel, die das Programm heute noch aufweist, zeigen, daß noch intensive Forschungen durchgeführt werden müssen, um unsere Kenntnisse über die Physiologie artikulatorischer Bewegungsvollzüge zu erhöhen. Diese erst bilden die Voraussetzung für eine in allen Belangen wirklichkeitsgetreue Modellierung. Es ist anzunehmen, daß die prinzipielle Möglichkeit, Sprechbewegungen computergrafisch zu modellieren, der Phonetik einen starken Impuls für diesbezügliche Forschungen verleiht.

7. LITERATUR

- /1/ MENZERATH, P. u. A. DE LACERDA: Koartikulation, Steuerung und Lautabgrenzung. Bonn 1934
- /2/ WÄNGLER, H.H.: Atlas deutscher Sprachlaute. Berlin, 7. Aufl. 1981
- /3/ Autorenkollektiv (Leitung: H. STÖTZER): Großes Wörterbuch der deutschen Aussprache. Leipzig 1982
- /4/ FANT, G.: Phonetik und Sprachforschung. In: Handbuch der Stimm- und Sprachheilkunde (LUCHSINGER/ARNOLD). Wien/New York 1970, S. 274
- /5/ LINDNER, G.: Grundlagen und Anwendung der Phonetik. Berlin 1981, S. 89 ff.
- /6/ RAUBER/KOPF: Lehrbuch und Atlas der Anatomie des Menschen. Bd. II, Leipzig 1955, S. 66
- /7/ MEHNERT, D.: Analyse und Synthese suprasegmentaler Intonationsstrukturen, ein Beitrag zur Optimierung technischer Sprachkommunikationssysteme. Diss.(B), Techn. Univ. Dresden, 1985, S. 98 ff.
- /8/ LAZICIUS, J.: Lehrbuch der Phonetik. Berlin 1961, S. 120 ff.
- /9/ LINDNER, G.: Der Sprechbewegungsablauf. Eine phonetische Studie des Deutschen. Berlin 1975, S. 97 ff.
- /10/ GABKA, K.: Einführung in das Studium der russischen Sprache. Bd. I Phonetik und Phonologie (E. WIEDE). Leipzig 1974, S. 85 - 95
- /11/ LINDNER, G.: Das Motogramm - ein Mittel zur Veranschaulichung artikulatorischer Verläufe. ZPSK 30 (1977), S. 535-543