

# Visuelle Animation der Artikulation als Therapiehilfe bei Sprechstörungen: Eine neurophonetische Begründung

## Visual Animation of Articulation as an Aid in Therapy for Speech Disorders: A Neurophonetic Rationale

**Autor**

**B. J. Kröger**

**Institut**

Klinik für Phoniatrie, Pädaudiologie und Kommunikationsstörungen Universitätsklinikum Aachen und RWTH Aachen (RWTH Aachen University) (Direktorin: Univ.-Prof. Dr. med. Ch. Neuschaefer – Rube)

### Schlüsselwörter

- Artikulation
- Lautartikulation
- Lautbilder
- visuelle Stimulation
- Therapie von Sprechstörungen
- neurophonetisches Modell der Sprachproduktion
- Sprachwahrnehmung

### Key words

- articulation
- speech sounds
- visual stimulation
- therapy of speech disorders
- neurophonetic model of speech production
- speech perception

### Bibliografie

DOI 10.1055/s-0029-1242738  
 Sprache · Stimme · Gehör  
 2009; 33: 179–185  
 © Georg Thieme  
 Verlag KG Stuttgart · New York  
 ISSN 0342-0477

### Korrespondenzadresse

**Prof. Dr. B. J. Kröger**  
 Klinik für Phoniatrie  
 Pädaudiologie und  
 Kommunikationsstörungen  
 Universitätsklinikum Aachen  
 und RWTH Aachen  
 (RWTH Aachen University)  
 Pauwelsstraße 30  
 52074 Aachen  
 bkroeger@ukaachen.de

### Zusammenfassung

▼  
 Darstellungen der Lautartikulation (Lautbilder) werden gerne als visuelle Hilfsmittel in der Therapie von Sprechstörungen eingesetzt. In diesem Artikel wird ein computerimplementiertes Programm zur Generierung von medio-sagittalen Lautbildern sowie artikulatorischen Bewegungsabläufen ganzer Silben und Wörter vorgestellt. Es wird erläutert, dass dieses Programm als visuelles Hilfsmittel oder visuelle Stimulation neben der Förderung der korrekten Produktion von Wörtern auch zur Förderung der taktilen, propriozeptiven und auditiven Wahrnehmung der Silben- oder Wortproduktion genutzt werden kann. Darüber hinaus wird anhand eines neurophonetischen Modells der Sprachproduktion, Sprachwahrnehmung und des frühen Spracherwerbs begründet, dass diese visuelle Stimulationsmethode aufgrund des engen Zusammenhangs zwischen motorischen, taktilen, propriozeptiven und visuellen Mustern der Silben- und Wortproduktion in der Therapie unterschiedlicher Sprechstörungen erfolgreich eingesetzt werden kann.

### Abstract

▼  
 Visual representations of the articulation of speech sounds (articulatory sound pictures) are widely used as a visual aid in the therapy for speech disorders. In this article, a computer-implemented algorithm for the generation of mediosagittal sound pictures as well as articulatory motion sequences of whole syllables and words is introduced. It is demonstrated that this programme can be used not only as a visual aid or visual stimulation method for the correct production of words but also as a stimulation method for developing or enhancing tactile, proprioceptive and auditory perception in syllable or word production. In addition, a neurophonetic model of speech production, perception and early speech acquisition – which is capable of explaining the close connection between motor, tactile, proprioceptive and visual patterns of the syllabic production and word production – is introduced in order to motivate the successful application of this visual stimulation method in therapy for different types of speech disorders.

### Lernziel

- ▶ Erläuterung von Arbeitsmodi und Einsatzmöglichkeiten eines Computerprogramms zur visuellen Animation von Artikulationsbewegungen in der Therapie von Sprechstörungen
- ▶ Erläuterung des Zusammenhanges zwischen Sprachproduktion, Sprachwahrnehmung und frühen Phasen des Spracherwerbs anhand eines neurophonetischen Modells
- ▶ Begründung visueller Stimulation in der Therapie von Sprechstörungen aus neurophonetischer Sicht

### Einleitung

▼  
**Rückkopplungsgestützte Hilfsmittel** Visuelle Hilfsmittel in der Therapie von Sprechstörungen können rückkopplungsgestützt sein oder nicht. Diese Rückkopplung kann akustisch oder artikulatorisch erfolgen. Bei akustischer Rückkopplung wird das eventuell fehlerhafte Sprachsignal des Patienten in Echtzeit in möglichst eingängiger Form visualisiert und kann dann mit der Visualisierung eines korrekt realisierten Sprachsignals verglichen werden [z.B. 1–4]. Bei artikulatorischer oder aerodynamischer Rückkopplung werden artikulatorische Parameter des Patienten (z.B. Zungenposition mittels der EMA-Methode

[5,6], Kontaktfläche zwischen Zunge und hartem Gaumen mittels EPG-Methode [7,8], Position des Gaumensegels mittels Nasopharyngoskop [9]) oder aerodynamische Daten des Patienten (z.B. nasaler vs. oraler Luftstrom [10–12]) während des Sprechens aufgezeichnet, in Echtzeit visualisiert und ebenfalls mit einer artikulatorischen Visualisierung eines korrekt realisierten Sprachsignals verglichen.

Die Problematik bei der akustischen Rückkopplung ist, dass ein hochwertiges Aufnahme-Mikrofon und eine hochwertige Signalvorverstärkung und Analog-nach-Digital-Wandlung notwendig ist. Dies setzt spezielles Equipment voraus. Seit einigen Jahren kann dieses Problem aber durch Nutzung von USB-Mikrofonen bzw. USB-Head-Sets gelöst werden, sodass man den akustischen Feedbackverfahren in der Therapie von Sprechstörungen zukünftig eine bessere Verbreitung prognostizieren kann. Problematischer ist die Situation hingegen bei artikulatorischen und aerodynamischen Feedbackverfahren. Hier kann auch zukünftig nicht auf Standard-Input-Geräte zurückgegriffen werden und die oft nur in kleinen Stückzahlen produzierten artikulatorischen und aerodynamischen Messgeräte sind kostenintensiv.

Keine Rückkopplungsmethode hat sich bis heute flächendeckend durchgesetzt, da eine spezielle und oft teure Ausrüstung notwendig ist.

**Nicht-rückkopplungsgestützte Hilfsmittel** Anders ist die Situation bei nicht-rückkopplungsgestützten visuellen Hilfsmitteln. Diese sind aus Nutzersicht wesentlich unproblematischer, da eben keine Verarbeitung von Sprachsignalen des Patienten stattfindet. Es ist lediglich ein handelsüblicher Computer nötig. Der Nachteil nicht-rückkopplungsgestützter visueller artikulatorischer Hilfsmittel ist allerdings, dass diese Hilfsmittel dem Patienten nur „Normartikulationen“ vorgeben, diese aber nicht in Beziehung zum aktuellen artikulatorischen Produktionsvermögen des Patienten setzen können. Nicht-rückkopplungsgestützte visuelle artikulatorische Hilfsmittel in der Therapie von Sprechstörungen sind Artikulationsmodelle, die visuelle Hilfen zur Lautbildung beispielsweise in Form von „artikulatorischen Lautbildern“ (mediosagittale Schnittbilder zur Veranschaulichung der Stellung insbesondere der Zunge) geben [13–24]. Diese Artikulationsmodelle basieren auf statischen und dynamischen Artikulationsdaten eines oder mehrerer Sprecher und erlauben die Visualisierung von Artikulationsstellungen einzelner Laute aber auch die Visualisierung von Artikulationsbewegungen, wie sie bei der Produktion von Silben und Wörtern auftreten.

Nicht-rückkopplungsgestützte Hilfsmittel können dem Patienten nur die normale Artikulation vorgeben, sie aber nicht in Bezug auf seine aktuellen Fähigkeiten setzen.

Es ist das Ziel dieses Beitrages, das 2-dimensionale Artikulationsmodell „SpeechTrainer“ nach Kröger [16,19] und seinen Einsatz als Hilfsmittel in der Therapie von Sprechstörungen vorzustellen. Darüber hinaus wird der Wirkmechanismus visueller Stimulation durch animierte Artikulationsbewegungen in der Therapie von Sprechstörungen vor dem Hintergrund eines neurophonetischen Modells der Sprachproduktion, Sprachwahrnehmung und des frühen Spracherwerbs diskutiert.

## Computermodell zur visuellen Animation der Artikulation

Basierend auf artikulatorischen Messdaten mehrerer Sprecher des Standarddeutschen [17,25,26] wurden ein 2D- und ein 3D-Computermodell zur visuellen Darstellung des Sprechtraktes entwickelt. Diese Modelle sind in der Lage, sowohl statische Lautbilder wie auch dynamische Bewegungsabläufe (Artikulationsbewegungen) zu simulieren und visualisieren [16,27]. Das 2D-Modell der Artikulation (Programm „SpeechTrainer“ [28]) wurde bereits erfolgreich in der Therapie von Sprechstörungen eingesetzt ([29,30], siehe unten „Einsatz des Computermodells in der Therapie von Sprechstörungen“).

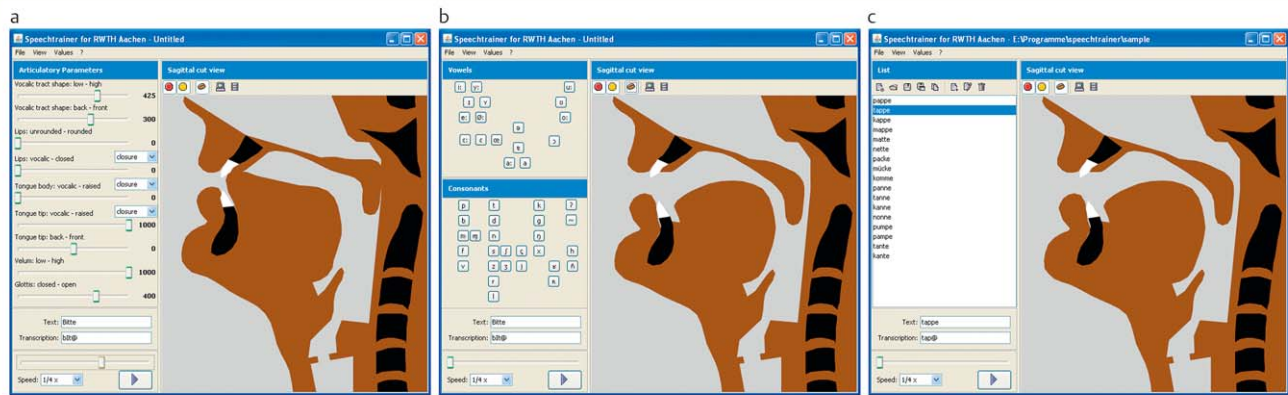
**Arbeitsmodi** Das Programm verfügt über die Arbeitsmodi *Artikulatoren*, *Laute* und *Wortliste*. Im Modus *Artikulatoren* erlaubt das Programm die freie Einstellung und Bewegung der Modellartikulatoren (Lippen, Zungenrücken, Zungenspitze, Gaumensegel, Glottis, s. **Abb. 1a**). Im Modus *Laute* erlaubt das Modell die Auswahl beliebiger Konsonanten oder Vokale des Standarddeutschen und zeigt die zugehörige Einstellung der Artikulatoren im Mediosagittalschnitt als Standbild (**Abb. 1b**). Im Modus *Wortliste* kann ein beliebiges Wort des Standarddeutschen orthografisch oder auch direkt in phonetischer Transkription eingegeben werden, oder aber aus einer bereits vom Therapeuten vorab erstellten Wortliste abgerufen werden (**Abb. 1c**). Dann kann der artikulatorische Bewegungsablauf dieses Wortes als Animation in Echtzeit oder in verschiedenen Zeitlupen-Stufen wiedergegeben werden. Der artikulatorische Ablauf vollzieht sich dabei nicht in abrupt-segmentaler Weise, d.h. indem ein statisches Lautbild an das nächste gereiht wird, sondern berücksichtigt koartikulatorische Prozesse [16,25] und erstellt somit die real auftretenden fließenden und von Laut zu Laut stetig ineinander übergehenden Bewegungen der Artikulatoren bei der Wort- und Satzproduktion.

**Benutzeroberfläche und Beispiel** **Abb. 1** zeigt die Oberfläche des Programms SpeechTrainer in allen drei Modi: *Artikulatoren*, *Laute* und *Wortliste*. **Abb. 2** zeigt Ausschnitte aus der Animation des Wortes „Kanu“. Diese Bildfolge kann allerdings nur sehr eingeschränkt einen Eindruck von der real im Modell möglichen visuellen Animation von Artikulationsbewegungen geben. Video-Beispiele zur visuellen Animation sind unter [31] zu finden. In der Therapie von Sprechstörungen können alle 3 Modi des Programms genutzt werden.

## Einsatz des Computermodells in der Therapie von Sprechstörungen

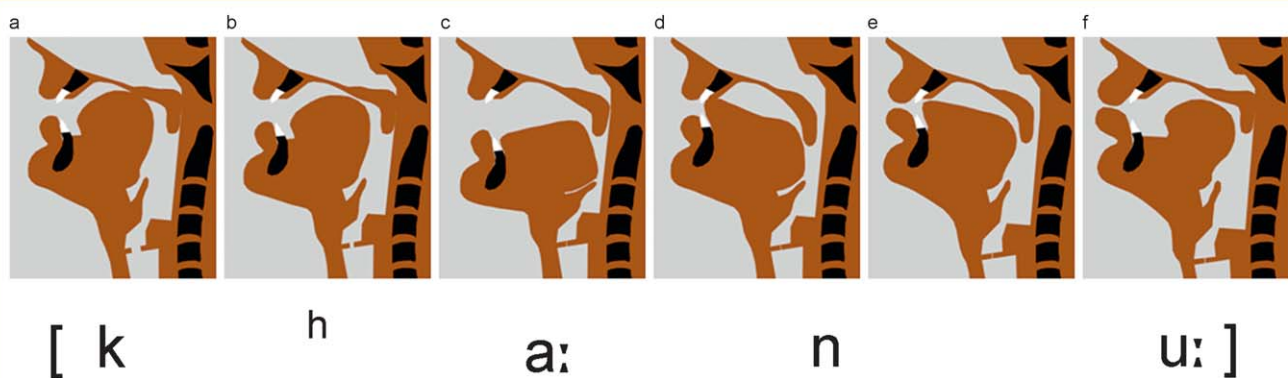
Bei verschiedenen Arten von Sprechstörungen z.B. bei kindlichen Artikulationsstörungen, bei fehlerhafter Artikulation aufgrund einer Hörschädigung, bei erworbenen neurogenen Sprechstörungen wie z.B. einer Sprechapraxie empfiehlt sich die begleitende Nutzung des visuellen Computermodells (SpeechTrainer) in verschiedenen Phasen des Therapieverlaufs [18,29,30].

**Förderung der taktilen Wahrnehmung** Der Therapeut stellt im Modell einen offenen Vokal ein (z.B. [a]) und stimuliert beim Patienten definierte Bereiche auf dessen Zungenspitze, Zungenrücken, am harten oder weichen Gaumen, bzw. auf Ober- oder Unterlippe. Der Patient muss den wahrgenommenen Ort dann



**Abb. 1** Darstellung des Hauptfensters des Programms „SpeechTrainer“ im Modus **a) Artikulatoren**, **b) Laute** und **c) Wortliste**. Rechts jeweils Bildschirmausschnitt für die Darstellung des Mediosagittalschnitts bzw. für die Darstellung der Animation der Artikulationsbewegungen. Links das Modus-bezogene Fenster. **a)** Fenster mit Schiebereglern für die Einstellung von 9 artikulatorischen Parametern. Von oben: Vokalische Formung tief bis hoch (z. B. [a] bis [i]); vokalische Formung hinten bis vorne (z. B. [u] bis [i]); Lippen: ungerundet bis gerundet (z. B. [i] bis [y]); Lippen: vokalisch bis konsonantisch geschlossen (z. B. Verschlussbildung für [p]); Zungenrücken: vokalisch bis konsonantisch gehoben (z. B. Verschlussbildung für [k]); Zungenspitze: vokalisch bis konsonantisch gehoben (z. B. Verschlussbildung für [t]); Zungenspitze gehoben: hinten bis vorne (z. B. [ʃ] bis [s]); Gaumensegel tief bis hoch (nasal bis nicht-nasal); Stimmritze: fest geschlossen bis geöffnet ([ʔ] über Phonation weiter über [h] bis

Atemstellung). Dazu Auswahlmöglichkeit der Verschlussbildung (Verschluss vs. kritische Enge zur Geräuschbildung vs. Verschluss mit lateraler Öffnung) für 3 artikulatorische Parameter. **b)** Fenster mit Klickbuttons für Lang- und Kurzvokale und reduzierte Vokale des Standard-Deutschen. Links unten: Konsonanten des Standard-Deutschen. Nach Anklicken erscheint das zugehörige Lautbild im rechten Fenster. **c)** Fenster der Wortliste, in dem Einzelwörter aktiviert werden können. Darunter: Darstellung des ausgewählten Wortes orthografisch und in phonetischer Transkription. Darunter: Fortschrittsbalken zum Ablauf der Animation eines eingegebenen oder aus der Liste ausgewählten Wortes (auch im Artikulatoren- und Lautfenster). Darunter: Einstellung der Abspielgeschwindigkeit (Zeitlupe) und daneben Startschaltfläche für die Ausführung der Animation für die Artikulation des Wortes (auch im Artikulatoren- und Lautfenster).



**Abb. 2** Mediosagittale Schnittbilder zu definierten Zeitpunkten des Artikulationsablaufes des Wortes „Kanu“ und phonetische Transkription. **a)** Beginn der oralen Verschlussbildung des [k], **b)** Verschlusslösung des [k] (Phase der Behauchung [h]), **c)** Mitte des Vokals [a:], **d)** Beginn der Verschlussbildung des [n], **e)** Verschlusslösung des [n], **f)** Mitte des Vokals [u:]. Anmerkung: Die auftretende Koartikulation ist insbesondere bei der Produktion des [n] gut zu erkennen. Zu Beginn der Verschlussbildung nehmen Lippen und Zungenrücken noch weitgehend die Form des vorhergehenden Vokals [a:] an, während die Stellung der Lippen und des Zungenrücken bei der Verschlusslösung bereits weitgehend durch den nachfolgenden Vokal [u:] bestimmt werden.

auf dem Bildschirm (► **Abb. 1**) zeigen. Diese Übung fördert die taktile Selbstwahrnehmung des Patienten und gleichzeitig die Vertrautheit mit dem mediosagittalen Schnittbild des Modells.

**Förderung der propriozeptiven Wahrnehmung** Der Therapeut variiert im Modus „Artikulatoren“ die Position der Zunge (vorne-hinten, hoch-tief), die Position der Zungenspitze (gehoben-neutral) oder die Position der Lippen (geöffnet-geschlossen, gerundet-ungerundet). Der Patient wird angehalten, diese vom Modell gezeigten Bewegungen stumm nachzuvollziehen und dabei die unterschiedlichen Stellungen der Artikulationsorgane

(propriozeptiv) wahrzunehmen. Schwierig gestaltet sich insbesondere die (bewusste) Bewegung des Gaumensegels und der Stimmritze. Für das Gaumensegel kann im Modus „Artikulatoren“ die Hebe- und Senkbewegung des Gaumensegels gezeigt werden. Parallel kann z. B. die Hebebewegung des Gaumensegels durch Lautabfolgen Nasal-Obstruent (also z. B. [ŋga]) geübt werden. Die Öffnungsweite der Stimmritze kann im Modus „Artikulatoren“ von festem Verschluss über normale Phonation, behauchte Phonation und Hauchstellung bis hin zur Atemstellung variiert werden. Parallel kann der Therapeut diese Modi dem Patienten vorstellen und dann mit ihm üben.

## Neurophonetisches Modell der Sprachproduktion, Sprachwahrnehmung und des frühen Spracherwerbs

**Prälinguistische Phase des Spracherwerbs** Die Struktur des neurophonetischen Modells (◉ **Abb. 3** [32]) ist während des frühen Spracherwerbs noch nicht vollständig entwickelt. Das Kleinkind ist zunächst nur in der Lage, motorische Pläne aufgrund zufällig aktivierter (vorsprachlicher) Artikulatorbewegungen (Lippenschließung/-öffnung, Zungenhebung/-senkung, sprachähnliche Phonation, ...) zu generieren. Dies kennzeichnet die prälinguistische Phase des Spracherwerbs ([36]: „Babbeln“). Ein motorischer Plan wird dabei in eine Folge primär-motorischer Bewegungszustände umgesetzt (motorische Programmierung) und führt über neuromuskuläre Aktivierungen zu definierten Artikulatorbewegungen (Sprechtraktmodell [27]). Das so gesteuerte Sprechtraktmodell definiert für jeden Zeitpunkt die Stellung (und Bewegung) aller Artikulatoren und das akustische Signal [38]. Aus der momentanen Stellung der Artikulatoren resultiert dann das artikulatorische und optische Signal, aus dem dann wiederum das somatosensorische Feedbacksignal (taktile und propriozeptive), das visuelle sowie das auditive Feedbacksignal gewonnen werden können (sensorische Rezeptoren und Vorverarbeitung). Die momentanen sensorischen Zustände werden dann auf Silben- oder Wortlänge als sensorische Feedback-Muster (Rückkopplungsmuster, RM, ◉ **Abb. 3**) gespeichert. Über die phonetische Karte (◉ **Abb. 3**) wird nun der zugrunde liegende motorische Plan mit den erhaltenen sensorischen Mustern (auditiv und somatosensorisch) verknüpft. Damit wird der motorische Plan zusammen mit den zugehörigen sensorischen Mustern gespeichert, sodass das Kleinkind nun in der Lage ist, einen entsprechenden motorischen Plan (motorisches Muster) anhand der gelernten (gespeicherten) sensorischen Muster erneut zu aktivieren. Während der prälinguistischen Phase des Spracherwerbs baut das Kleinkind somit bereits erste Vorstellungen auf, wie ein bestimmter motorischer Plan „klingen“ wird (gelerntes auditives Muster) und wie sich die Produktion bzw. Artikulation dieses motorischen Plans „anfühlen“ wird (gelerntes somatosensorisches Muster). Damit ist das Kleinkind dann bereits in der Lage, externe sensorische Muster von Kommunikationspartnern (z. B. der Mutter) zu imitieren.

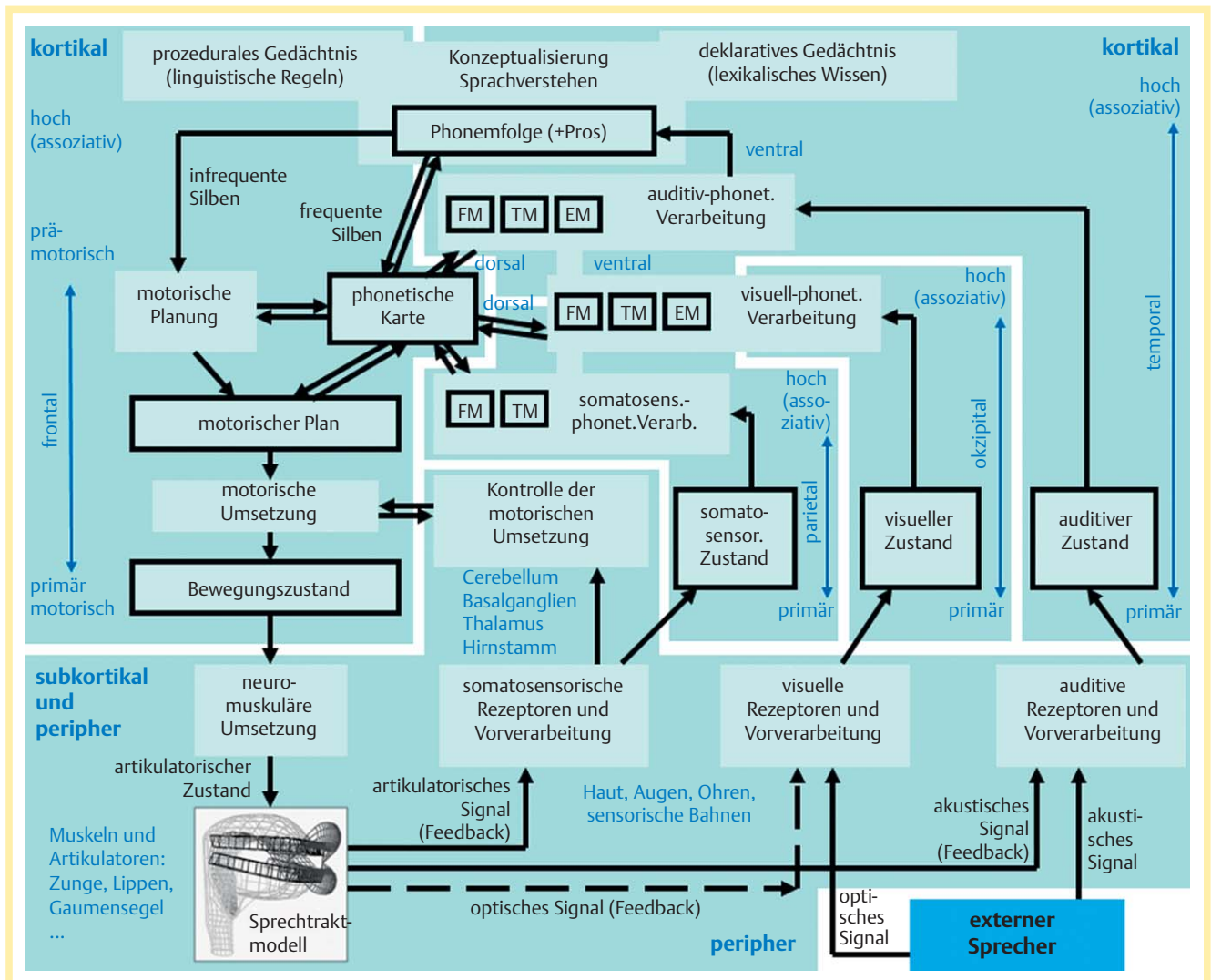
**Imitationsphase des Spracherwerbs** Es beginnt jetzt die Imitationsphase des Spracherwerbs. Ein externer Sprecher spricht ein Wort und aktiviert damit beim lernenden Kleinkind (d. h. im Modell) ein externes auditorisches Muster (EM). Dies wird mit bereits früher gelernten sensorischen Mustern verglichen und es wird das nächstliegende vorab gelernte sensorische Muster und der zugehörige motorische Plan via der phonetischen Karte aktiviert. Das sensorische Resultat aufgrund der Produktion dieses Plans (sensorisches Feedbackmuster, FM) wird dann mit dem externen sensorischen Muster (EM) im auditiv-phonetischen Verarbeitungsmodul verglichen. Ist die Imitation des Wortes oder einer Silbe des Wortes noch nicht befriedigend, kann der motorische Plan abgewandelt (iterativ verbessert) und ein erneuter Imitationsversuch gestartet werden. Fällt das Imitationsergebnis zufriedenstellend aus, wird der aktuelle motorische Plan der Silbe oder des Wortes zusammen mit seinem sensorischen Muster (auditiv und somatosensorisch) und seinem phone-

mischen Aktivierungsmuster (Phonemfolge) über die phonetische Karte verknüpft und gespeichert. Am Ende der Imitationsphase, die wiederum mit der Phase des Lexikonaufbaus zeitlich überlappt, ist das Kind dann in der Lage, alle in der Muttersprache häufig auftretenden Silben (frequente Silben) phonemisch und motorisch zu aktivieren und zu produzieren (Konzept des mentalen Silbenspeichers [39,40]). Gleichzeitig werden bei der Produktion über die phonetische Karte neben dem motorischen Plan auch die gelernten (trainierten) sensorischen Muster (TM) aktiviert. Das bedeutet, dass der Sprecher bereits vor der Produktion einer häufig auftretenden Silbe weiß, wie diese klingt (auditives Muster), wie die zugehörige Mundbewegung aussieht (visuelles Muster), und wie sich die Produktion „anfühlt“ (somatosensorisches Muster). Zudem kann bei der Wahrnehmung häufig auftretender (und damit gelernter) Silben neben den sensorischen Mustern über die phonetische Karte auch der motorische Plan der Silbe aktiviert werden (dorsaler Pfad der Sprachwahrnehmung und Spiegelneuronenhypothese [41]).

**Förderung der auditiven Lautwahrnehmung** Nur im Falle von Hörstörungen oder eingeschränkter auditiver Lautwahrnehmung spricht der Therapeut einen perzeptiv (noch nicht produktiv) zu erlernenden Laut isoliert oder im Silbenkontext vor und bietet gleichzeitig die artikulatorische Information dazu visuell an (Standbild oder Animation mittels des Programms). Der Patient wird angehalten, sowohl auf den auditiven wie auf den visuellen Eindruck zu achten. Es kann angenommen werden, dass dies dem Patienten hilft, das akustisch wahrgenommene Klangbild des Lautes direkt mit einem visuellen Bild der Artikulation und damit auch mit der entsprechenden somatosensorischen (d. h. taktilen und propriozeptiven) Wahrnehmung zu verknüpfen. Wird der Laut audio-visuell stabil erkannt, sollte im Verlauf der Therapie das visuelle Signal immer seltener dargeboten werden, um die letztlich wichtige auditive Identifikation des wahrnehmungs-problematischen Lautes zu ermöglichen. Während beispielsweise bei Vokalen und Frikativlauten durchaus Standbilder (Einzellautproduktion) genutzt werden können, ist bei der Wahrnehmung insbesondere von Plosivlauten und Nasallauten die Artikulationsbewegung und ihr perzeptives Korrelat (Formanttransition) wichtig. Hier ist daher die Einbettung des Lautes in CV- oder VC-Silben und damit neben den Standbildern die Nutzung der Animationen sinnvoll.

**Förderung der Laut-, Silben- und Wortartikulation** Wie bei der Förderung der Lautwahrnehmung spricht der Therapeut auch hier den Laut, die Silbe oder das Wort mit, während er die visuelle Information des Modells (Standbild oder Animation) zeitgleich darbietet. Der Patient wird dann aufgefordert, den Laut, die Silbe oder das Wort wiederholend mit dem Ablauf der Animation selbst zu produzieren. Während bei der Anbahnung von Einzellauten Standbilder bereits sehr hilfreich sein können, ist gerade im Fall komplexer motorischer Störungen – wie z. B. der Sprechapraxie – die Darbietung der Animation von Artikulationsbewegungen und damit die Darbietung der Lautverbindungen und der koartikulatorischen Abläufe sehr wichtig. Patienten mit diesen Störungsbildern können nämlich oft Einzellaute bilden, sind aber nicht in der Lage, diese zu Silben und Wörtern zusammenzufügen.





**Abb. 3** Aufbau des neurophonetischen Modells der Sprachproduktion, Sprachwahrnehmung und des frühen Spracherwerbs [32]. Umrandete Kästen kennzeichnen neuronale Karten, die jeweils bestimmte motorische, sensorische, phonetische oder phonemische Zustände repräsentieren. Nicht umrandete Kästen kennzeichnen komplexe Prozessmodule. Einfache Pfeile repräsentieren neuronale Projektionen, z. B. für die einfache Weiterleitung von Information. Doppelpfeile kennzeichnen komplexe neuronale Assoziationen, z. B. für die Aktivierung von komplexen motorischen oder sensorischen neuronalen Mustern aus einfachen phonemischen oder phonetischen Mustern. Gestrichelte Linie: ein optisches Feedbacksignal entsteht nur, wenn sich der Sprecher beim Sprechen im Spiegel betrachtet. Die neuronalen Karten innerhalb der sensorisch-

phonetischen Verarbeitungsmodulen kennzeichnen trainierte (also erlernte) sensorische Muster (TM), aktuell produzierte Feedbackmuster (FM) und extern produzierte (also von dem Kommunikationspartner produzierte) sensorische Muster (EM). „Pros“ steht für prosodische Information auf der phonemischen Ebene. Der Bewegungszustand und die sensorischen Zustände repräsentieren jeweils einen einzelnen Zeitpunkt (neuronale Karten im primären frontalen, temporalen, parietalen und okzipitalen Kortexareal), während der motorische Plan und alle sensorischen Muster (FM, TM, EM) den zeitlichen Ablauf einer ganzen Silbe repräsentieren (neuronale Karten in höheren uni- oder heteromodalen kortikalen Assoziationsarealen, die in enger Vernetzung mit Kurz- bzw. Langzeitgedächtnis stehen).

### Begründung des Programmeinsatzes als Therapiehilfe

Das neurophonetische Modell der Sprachproduktion, Sprachwahrnehmung und des frühen Spracherwerbs nach Kröger ([32], siehe auch Box „Neurophonetisches Modell der Sprachproduktion, Sprachwahrnehmung und des frühen Spracherwerbs“), wie auch das Modell von Guenther [33] verdeutlichen den engen Zusammenhang von auditiver und propriozeptiver Wahrnehmung (Sensorik) mit der Bewegungssteuerung und -ausführung (Motorik) beim Sprechen.

**Prälinguistische Phase des Spracherwerbs** Während des frühen Spracherwerbs (ca. die ersten 2 Lebensjahre) erlernt das

Kleinkind bereits sprachspezifische motorische Pläne zur Realisierung von Silben und auch von ersten Wörtern und die zugehörigen auditiven und somatosensorischen Muster. Während der prälinguistischen Phase des Spracherwerbs („Babbeln“ oder „Lallen“) kombiniert das Kleinkind freie (d. h. noch nicht sprachspezifische) Artikulationsbewegungen mit sprachähnlicher Phonation [34]. Dadurch sammelt es bereits Wissen um den Zusammenhang von motorischen Plänen eines artikulatorischen Ablaufes mit deren somatosensorischen Korrelaten (d. h. wie sich die Produktion dieses artikulatorischen Ablaufes taktil und propriozeptiv „anfühlt“) und mit deren auditiven Korrelaten (d. h. wie sich die eigene Produktion dieses artikulatorischen Ablaufes „anhört“). Die Prozesse des in dieser prälinguistischen Phase des

Spracherwerbs stattfindende Wissenszuwachs wurde von Kröger et al. [32] mittels künstlicher neuronaler Netze simuliert.

**Linguistische Phase des Spracherwerbs** Das in der prälinguistischen Phase des Spracherwerbs gesammelte Wissen ist dann die Basis für die ab dem 2. Lebensjahr parallel einsetzende linguistische (sprachspezifische) Phase des Spracherwerbs. In dieser 2. Phase ist zunächst die Imitation des sprachlichen Verhaltens der nächsten Bezugspersonen (z. B. Mutter) wichtig. Die Basis für die Imitation ist das vom Kommunikationspartner produzierte akustische Sprachsignal, aber zusätzlich auch das Mundbild, d. h. die Information zu den sprechbedingten Bewegungen des Unterkiefers und der Lippen. Dass das Kleinkind nun überhaupt in der Lage ist, diese akustischen (und visuellen) Muster erfolgreich zu imitieren, d. h. akustisch zu reproduzieren, setzt genau das während der prälinguistischen Phase erworbene Wissen zum Zusammenhang zwischen motorischen Plänen von artikulatorischen Abläufen und deren somatosensorischen und auditiven Konsequenzen voraus. Das Imitieren der Sprechmuster des Kommunikationspartners durch das Kleinkind führt zu immer perfekteren Artikulationen erster Silben und Wörter der Muttersprache. Neben den somatosensorischen Mustern werden aber auch visuelle Muster (Mundbewegungen der Bezugspersonen) verarbeitet, die nachweislich motorische und somatosensorische Information integrieren können [35]. So lernt das Kleinkind während der Imitationsphase auch die Verknüpfung von visuellen Mustern mit motorischen Plänen von artikulatorischen Bewegungsmustern und deren sensomotorischen Konsequenzen. Nach erfolgreichem Imitationstraining ist das Kleinkind dann in der Lage, motorische Pläne zu den häufig trainierten Silben der Muttersprache auf der Ebene des mentalen Silbenspeichers abzurufen und damit ohne aufwändige auditive Kontrolle richtig zu produzieren. Das Kleinkind hat das artikulatorische Muster dieser häufigen Silben während des Spracherwerbs so oft produziert, dass dieser motorische Plan und das zugehörige Artikulationsmuster als „überlernt“ bezeichnet werden können [36].

**Sprechstörungen** Bei vielen Sprechstörungen liegt mehr als eine isolierte neuromuskuläre Fehlfunktion bei der Sprachproduktion vor. Es kann angenommen werden, dass auch der Zugriff auf motorische Pläne bzw. der Zugriff auf die Verknüpfung von motorischen Plänen und sensorischen Mustern gestört ist [37]. Insbesondere beim Erlernen von Silben und Wörtern ist aber die Aktivierung der auditiven und der propriozeptiven und taktilen Muster der zu realisierenden Silbe (des zu realisierenden Wortes) wichtig, um dem Kind (bzw. dem Patienten) die Lautziele anzugeben, die während des Sprechens realisiert werden müssen. Hier soll das Programm zur Visualisierung von Artikulationsbewegungen eingreifen. Aufgrund des oben dargestellten Zusammenhanges von Sprechmotorik und Sprechsensorik kann folgendes angenommen werden:

- ▶ Kann keine oder nur eine verminderte Zuordnung zwischen auditiven Mustern und motorischen Plänen, die dieses Muster realisieren, aktiviert werden, so ist es sinnvoll, den artikulatorischen Ablauf visuell neben dem auditiven Signal anzubieten, um genau diese Zuordnung wieder zu stärken.

- ▶ Kann kein somatosensorisches Muster zu einer zu produzierenden Silbe aktiviert werden, so ermöglicht die visuelle Darbietung des Artikulationsablaufes im Anschluss an die oben beschriebenen Übungen zur taktil-propriozeptiven Stimulation eine Hilfestellung insbesondere zur Aktivierung entsprechender somatosensorischer Muster.

Bei der Produktion und Wahrnehmung häufig auftretender, damit gut gelernter und hinsichtlich der Produktion automatisierter Silben (frequente Silben), werden neben dem trainierten motorischen Plan auch das gelernte taktile, propriozeptive, auditive, und visuelle Muster der Silbe aktiviert.

## Diskussion und Ausblick

Zwar vollzieht sich der normale Prozess des Spracherwerbs vor allem über den auditiven Kanal, aber das hier vorgestellte neurophonetische Modell der Sprachproduktion, Sprachwahrnehmung und des frühen Spracherwerbs illustriert die Wichtigkeit aller sensorischen Kanäle (auditiv, somatosensorisch und visuell) bei der Produktion, Wahrnehmung und beim frühen Erwerb gesprochener Sprache. Insbesondere kann der somatosensorische Kanal und dadurch indirekt die korrekte Artikulation von Lauten, Silben und Wörtern durch visuelle Stimulation in Form von animierten Artikulationsbewegungen aktiviert werden. Allerdings gibt das hier vorgestellte Computerprogramm zur Visualisierung von Artikulationsbewegungen derzeit noch kein Feedback hinsichtlich der Korrektheit bei der Nachahmung eines visuell dargebotenen Artikulationsablaufes. Dies soll in den nächsten Jahren über die Einbeziehung von akustischen und optischen Signalen der Patienten realisiert werden.

## Hinweis

Diese Arbeit wurde gefördert mit Mitteln der Deutschen Forschungsgemeinschaft (DFG) Projekt Nr. Kr 1439/13-1 und Projekt Nr. Kr 1439/15-1.

## Fazit

Es wurde ein Computermodell zur visuellen Animation von Sprechbewegungen [28] vorgestellt, das in der Therapie von Sprechstörungen genutzt werden kann. Es fördert die taktile und propriozeptive Wahrnehmung, die Laut-, Silben- und Wortartikulation und im Falle eingeschränkter auditiver Lautwahrnehmung auch die korrekte auditive Lautwahrnehmung.

## Zur Person



Prof. Dr. Bernd J. Kröger absolvierte von 1979 bis 1985 sein Physikstudium an der Universität Münster (Dipl.-Phys.), promovierte 1989 und habilitierte sich 1997 an der Universität zu Köln im Fach Phonetik bei Univ.-Prof. Dr. Georg Heike (Direktor des Instituts für Phonetik der Universität zu Köln). Bis 1998 arbeitete er am Institut für Phonetik der Universität zu Köln als wissenschaftlicher

Assistent mit dem Schwerpunkt artikulatorische und akustische Phonetik. 1999 bis 2001 war er als Gastprofessor am Institut für Deutsche Sprache und Linguistik der Humboldt-Universität zu Berlin und am Zentrum für Allgemeine Sprachwissenschaft (ZAS) in Berlin tätig. Seit 2001 ist er wissenschaftlicher Angestellter an der Klinik für Phoniatrie, Pädaudiologie und Kommunikationsstörungen des Universitätsklinikums Aachen (Direktorin: Univ.-Prof. Dr. med. Christiane Neuschaefer-Rube) und außerplanmäßiger Professor an der medizinischen Fakultät der RWTH Aachen.

## Literatur

- 1 Öster AM. Teaching speech skills to deaf children by computer-based speech training. *STL-Quarterly Progress and Status Report* 1995; 36 (4): 67–75
- 2 Vicsi K, Hacki T. CoKo – Computergestützter Sprechkorrektor mit audiovisueller Selbstkontrolle für artikulationsgestörte und hörbehinderte Kinder. *Sprache Stimme Gehör* 1996; 20: 141–149
- 3 Vicsi K, Csatari F, Bakcsi Zs et al. Distance score evaluation of the visualised speech spectra at audio-visual articulation training. *Proceedings of EUROSPEECH* 1999, Budapest, Hungary; 1999; 1911–1914
- 4 *Better Accent Tutor*. 2009; <http://www.betteraccent.com/>
- 5 Katz WF, Bharadwaj SV, Carstens B. Electromagnetic articulography treatment for an adult with Broca's aphasia and apraxia of speech. *Journal of Speech, Language, and Hearing Research* 1999; 42: 1355–1366
- 6 Goozee J, Murdoch B, Ozanne A et al. Lingual kinematics and coordination in speech disordered children exhibiting differentiated versus undifferentiated lingual gestures. *Journal of Language and Communication Disorders* 2007; 42: 703–724
- 7 Wrench A, Gibbon F, McNeill AM et al. An EPG therapy protocol for remediation and assessment of articulation disorders. In *Proceedings of the 7th International Conference on Spoken Language Processing ICSLP02*, Denver, Colorado, USA; 2002; 965–968
- 8 Gibbon FE, McNeill AM, Wood SE et al. Changes in linguopalatal contact patterns during therapy for velar fronting in a 10-year-old with Down's syndrome. *International Journal of Language and Communication Disorders* 2003; 38: 47–64
- 9 Brunner M, Stellzig-Eisenhauer A, Pröschel U et al. The effect of nasopharyngoscopic biofeedback in patients with cleft palate and velopharyngeal dysfunction. *The Cleft Palate-Craniofacial Journal* 2005; 42: 649–657
- 10 Ruscello DM. Visual feedback in treatment of residual phonological disorders. *Journal of Communication Disorders* 1995; 28: 279–302
- 11 Main A. Notes and discussion instrumental assessment and treatment of hypernasality, following maxillofacial surgery, using SNORS: a single case study. *International Journal of Language and Communication Disorders* 1999; 34: 223–238
- 12 van Lierde K, Claeys S, de Bodd M et al. Outcome of laryngeal and velopharyngeal biofeedback treatment in children and young adults: A pilot study. *Journal of Voice* 2004; 18: 97–106
- 13 Badin P, Tarabalka Y, Elisei F et al. Can you “read tongue movements”? *Proceedings of Interspeech* 2008. Brisbane, Queensland, Australia; 2008; 2635–2638
- 14 Engwall O, Bälter O, Öster AM et al. Designing the user interface of the computer-based speech training system ARTUR based on early user tests. *Journal of Behaviour and Information Technology* 2006; 25: 353–365
- 15 Engwall O, Bälter O. Pronunciation feedback from real and virtual language teachers. *Journal of Computer Assisted Language Learning* 2007; 20: 235–262
- 16 Kröger BJ. Ein visuelles Modell der Artikulation. *Laryngo-Rhino-Otol* 2003; 82: 402–407
- 17 Kröger BJ, Hoole P, Sader R et al. MRT-Sequenzen als Datenbasis eines visuellen Artikulationsmodells. *HNO* 2004; 52: 837–843
- 18 Kröger BJ, Gotto J, Albert S et al. A visual articulatory model and its application to therapy of speech disorders: a pilot study In: S Fuchs, P Perrier, B Pompino-Marschall, Hrsg. *Speech production and perception: Experimental analyses and models*. ZAS Papers in Linguistics; 2005; 40: 79–94
- 19 Kröger BJ, Graf-Bortscheller V, Lowit A. Two- and three-dimensional visual articulatory models for pronunciation training and for treatment of speech disorders. *Proceedings of Interspeech* 2008. Brisbane, Queensland, Australia; 2008; 2639–2642
- 20 Massaro DW. *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle*. Verlagsort? MIT Press; 1998
- 21 Massaro DW. A computer-animated tutor for spoken and written language learning. *Proceedings of the 5th International Conference on Multimodal Interfaces*. Vancouver, British Columbia, Canada; 2003; 172–175
- 22 Massaro DW. The psychology and technology of talking heads: Applications in language learning In: van Kuppevelt J, Dybkjær L, Bernsen NO, eds. *Advances in Natural Multimodal Dialogue Systems*, Vol. 30. Heidelberg: Springer; 2005; 183–214
- 23 Massaro DW, Bigler S, Chen T et al. Pronunciation training: the role of eye and ear. *Proceedings of Interspeech* 2008. Brisbane, Queensland, Australia; 2008; 2623–2626
- 24 Massaro DW, Liu Y, Chen TH et al. A multilingual embodied conversational agent for tutoring speech and language learning. *Proceedings of Interspeech* 2006. Pittsburgh, PA, USA; 2006; 825–828
- 25 Kröger BJ. A gestural production model and its application to reduction in German. *Phonetica* 1993; 50: 213–233
- 26 Kröger BJ, Schröder G, Opgen-Rhein C. A gesture-based dynamic model describing articulatory movement data. *Journal of the Acoustical Society of America* 1995; 98: 1878–1889
- 27 Kröger BJ, Birkholz P. A gesture-based concept for speech movement control in articulatory speech synthesis In: Esposito A, Faundez-Zanuy M, Keller E, Marinaro M, eds. *Verbal and Nonverbal Communication Behaviours*. LNAI 4775. Berlin: Springer; 2007; 174–189
- 28 Kröger BJ. *SpeechTrainer: Visualisierung von Artikulationsbewegungen* 2009; <http://www.speechtrainer.de>
- 29 Funk J, Montanus S, Kröger BJ. Therapie von neurogenen und kindlichen Sprechstörungen mit dem Programm „SpeechTrainer“. *Forum Logopädie* 2006; 20: 6–13
- 30 Kendziorra I. *Artikulationstraining bei hörgeschädigten Kindern unter Einbeziehung des SpeechTrainers*. Diplomarbeit im Studiengang Lehr- und Forschungslogopädie. RWTH Aachen; 2006
- 31 Kröger BJ. Videos von Artikulationsbewegungen, generiert mittels des Modells SpeechTrainer 2009; <http://youtube.com/speechtrainer>
- 32 Kröger BJ, Kannampuzha J, Neuschaefer-Rube C. Towards a neurocomputational model of speech production and perception. *Speech Communication* 2009; 51: 793–809
- 33 Guenther FH. Cortical interaction underlying the production of speech sounds. *Journal of Communication Disorders* 2006; 39: 350–365
- 34 Oller DK, Eilers RE, Neal AR et al. Precursors to speech in infancy: the prediction of speech and language disorders. *Journal of Communication Disorders* 1999; 32: 223–245
- 35 Kröger BJ, Kannampuzha J. A neurofunctional model of speech production including aspects of auditory and audio-visual speech perception. *Proceedings of AVSP*. Moreton Island, Queensland, Australia; 2008; 83–88 Siehe auch <http://www.speechtrainer.eu>
- 36 Kröger BJ, Birkholz P, Neuschaefer-Rube C. Ein neuronales Modell zur sensorimotorischen Entwicklung des Sprechens. *Laryngo-Rhino-Otol* 2007; 86: 365–370
- 37 Kent RD. Research on motor speech control and its disorders: a review and prospective. *Journal of Communication Disorders* 2000; 33: 391–428
- 38 Birkholz P, Jackel D, Kröger BJ. Simulation of losses due to turbulence in the time-varying vocal system. *IEEE Transactions on Audio, Speech, and Language Processing* 2007; 15: 1218–1225
- 39 Levelt WJM, Wheeldon L. Do speakers have access to a mental syllabary? *Cognition* 1994; 50: 239–269
- 40 Indefrey P, Levelt WJM. The spatial and temporal signatures of word production components. *Cognition* 2004; 92: 101–144
- 41 Hickok G, Poeppel D. Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences* 2007; 4: 131–138