

# Eine komparative Studie artikulatorischer Rekonstruktion anhand akustischer Daten

Christian Geng,

Zentrum für Allgemeine Sprachwissenschaft, Berlin; Email: [geng@zas.gwz-berlin.de](mailto:geng@zas.gwz-berlin.de)

Bernd Pompino-Marschall,

Humboldt-Universität zu Berlin; Email: [bernd.pompino-marschall@rz.hu-berlin.de](mailto:bernd.pompino-marschall@rz.hu-berlin.de)

Ralf Winkler,

Institut für Kommunikationswissenschaften, TU Berlin; Email: [winkler@kgw.tu-berlin.de](mailto:winkler@kgw.tu-berlin.de)

## Einleitung

In der phonetischen Forschung trifft man immer wieder auf Fragestellungen, die beim gegenwärtigen Stand der Datensituation nicht beantwortbar sind. Ein solches Beispiel stellen die palatalen Plosive des Tschechischen dar. Die Besonderheit des tschechischen Phoneminventar besteht im Vorkommen palataler Plosive, die sich von den velaren Plosiven durch eine Vorverlagerung der Artikulationsstelle unterscheiden<sup>[1]</sup>. Aufgrund der spärlichen Datenlage wollten wir der Fragestellung nachgehen, inwieweit artikulatorische Inversionsmethoden die für die palatalen Verschlusslösungen relativ langen Engebildungen rekonstruieren können.

Dazu sollen selektiv Ansätze der Inversion von akustischer in artikulatorische Information empirisch verglichen werden. Diese lassen sich grob und ohne Anspruch auf Vollständigkeit folgendermaßen gliedern: LPC-Ansatz, Sortier- und Codebook-techniken sowie Stochastische Optimierung. Der LPC-Ansatz nutzt die Tatsache aus, dass das Filtern eines verlustfreien Röhrenmodells des Vokaltraktes dem optimalen inversen Filtern des Sprachsignals mit akkuraten Randbedingungen an Glottis und Lippen gleichkommt<sup>[2]</sup>. Sortiermethoden<sup>[3]</sup> sampeln die artikulatorischen Parameter eines beliebigen artikulatorischen Modells und konstruieren Tabellen von Vokaltraktgeometrien und damit assoziierten akustischen Repräsentationen, im Normalfall Formantfrequenzen. Ein anderes, dem Ansatz von Atal et. al. verwandtes Modell wurde in einem neueren Aufsatz von Story & Titze (1998) beschrieben<sup>[4]</sup>. Sie soll hier etwas genauer beschrieben werden, da sie von uns als Rekonstruktionsmodell benutzt wurde. In ihrem Modell werden mit Magnetresonanztomographie gemessene Areafunktionen auf eine konstante Anzahl röhrenförmiger Sektionen interpoliert. Von jeder individuellen Areafunktion wird eine neutrale, schwa-ähnliche Areafunktion, die sich als der Mittelwert der empirischen Areafunktionen berechnet, subtrahiert. Die daraus resultierende Matrix, welche Verwerfungen der Areafunktionen aus der Neutrallage beschreibt, wird einer Hauptkomponentenanalyse unterworfen und die beiden ersten Komponenten werden zurückbehalten. Anschließend wird für jeden der beiden Komponenten Mittelwert(m) und Standardabweichung(SD) bestimmt und der Bereich zwischen  $m+2SD$  und  $m-2SD$  in 50 äquidistante Intervalle unterteilt. Durch vollständige Kombination der resultierenden Faktorwerte entstehen 2500 synthetische Areafunktionen und durch Konversion der Areafunktion in spektrale Information und anschließendes Peakpicking erhält man eine Tabelle mit Faktorkoeffizienten und dazugehörigen F1/F2-Paaren. In einem nächsten Schritt wird eine data-gridding und Interpolationstechnik angewandt, mit deren Hilfe gegebene F1/F2-Paare mit Areafunktionen assoziiert werden. Abschließend sollen noch nonlineare Optimierungstechniken genannt werden. Einer dieser Ansätze, Simulated Annealing, ist als Erweiterung des klassischen Metropolis-Algorithmus zu verstehen<sup>[5]</sup>. Simu-

lated Annealing<sup>[6]</sup> ist eine stochastische Optimierungstechnik, mit deren Hilfe willkürliche Grade an Nonlinearität, Diskontinuität und Stochastizität verarbeitbar sein sollten. Die von uns benutzte Variante nennt sich Adaptive Simulated Annealing und stellt eine flexiblere Variante des ‚klassischen‘ Simulated Annealing dar<sup>[7]</sup>. Einen weiteren Kontrast stellt die Wahl des Ausgangsdatenmaterials dar: Areafunktionen als Repräsentationen der gesamten Ansatzrohrgeometrie oder deren Partionierung in einzelne Artikulatoräquivalente, aus deren Verhalten die Ansatzrohrgeometrie abgeleitet wird.

## Methoden

### Artikulatorische Datenerhebung

Zunächst wurde die dreidimensionale Formung des Ansatzrohrs während der Produktion der Vokale des Deutschen eines Sprechers (JD) mittels Magnetresonanztomographie (MRT) vermessen. Im folgenden wurden für jeden Vokal je nach Gesamtlänge des Ansatzrohrs zwischen 39 und 51 zur Luftströmungsrichtung orthogonale Schnittflächen durch das Volumen des Vokaltrakts errechnet. Aus diesen können direkt Areafunktionen als Funktion des aktuellen Abstands zur Glottis ermittelt werden. Das einfache artikulatorische Modell besteht in einer Hauptkomponentenanalyse der Abweichungen von der durchschnittlichen Areafunktion. Die Ermittlung der Übertragungsfunktion (128 Punkte, 0 ... 10 kHz) erfolgte mittels des Modells von Sondhi und Schroeter (1987<sup>[8]</sup>). Die Ermittlung der Formantfrequenzen wurde mittels Peak-Picking und parabolischer Interpolation realisiert, zur Bandbreitenextraktion wurde ein binärer Suchalgorithmus verwendet. Als artikulatorbasiertes Modell diente das 7-Parameter-Modell von Maeda<sup>[9]</sup> mit den Parametern Kieferposition, Zungenlage, Zungenform, Zungenspitze, Lippenöffnung, Lippenstülpung und Larynxhöhe.

### Akustische Datenerhebung

Das akustische Ausgangsmaterial bestand in den stimmhaften und stimmlosen palatalen Plosiven des Tschechischen in silbeninitialer Position, wie sie im Handbuch der International Phonetic Association beschrieben sind<sup>[10]</sup>. Diese Aufnahmen wurden lowpassgefiltert und mit 11kHz neu abgetastet. Daraus wurden Formantracks mit der Sprachsignalverarbeitungssoftware PRAAT extrahiert.

## Ergebnisse

Im folgenden Abschnitt erfolgt eine aus Platzgründen eher komprimierte Darstellung der Ergebnisse: Die mit dem Modell von Story & Titze gewonnenen Areafunktionen sind als realisationsphonetisch nicht plausibel zu werten. Zwar kann man den Konstriktionsort als befriedigend rekonstruiert ansehen, jedoch konnten die Form und Länge der Konstriktion nicht

physiologisch realistisch extrahiert werden. Die Vorteile dieser Methode sind darin zu sehen, daß ohne nennenswerten Mehraufwand Trajektorien generierbar sind und dass die Passung zwischen Areafunktionen und (allerdings nur den ersten beiden) Formanten a priori perfekt ist.

Dagegen scheitert der Optimierungsansatz bei den Areafunktionen. Die Rekonstruktionen waren nicht eindeutig, d.h. wiederholte Optimierungen, auch mit unterschiedlicher Wahl der Startparameter, konvergierten nicht zu einem stabilen Optimum. Die naheliegendste Hypothese, dieses Verhalten zu erklären besteht in der mangelnden ‚Topologizität‘ der Hauptkomponentenanalyse, d. h. ähnlichen Werten im Faktorenraum entsprechen nicht unbedingt ähnliche Areafunktionen. Umgekehrt lassen sich zu einer gegebenen Areafunktion mit –wie hier– vorgegebenen akustischen Eigenschaften Faktorkoeffizienten ermitteln, welche in unterschiedlichen Regionen des Faktorraumes liegen: Um diese Annahme zu überprüfen, wurde eine Tabelle generiert, die neben den ersten drei Hauptkomponenten die assoziierten extrahierten Formanten enthielt. Die Koeffizientenpaare mit dem geringsten prozentualen Fehlern in den ersten drei Formanten wurden zurückbehalten und einer Clusteranalyse unterworfen. Es zeigte sich, dass diese ‚guten‘ Faktorkoeffizienten distinkte Regionen im Faktorenraum besetzen. Entsprechend kann das Versagen des Optimierungsansatzes tatsächlich auf eine unwünschenswerte Abbildung der Areafunktionen in den Faktorraum zurückgeführt werden. Im Gegensatz dazu erwies sich das Modell von Maeda als dem Optimierungsansatz zugänglich: Ohne die Parametervariation allzu sehr einschränken zu müssen, lieferte der Ansatz befriedigende Ergebnisse sowohl im Hinblick auf die Passung mit den zu rekonstruierenden Formanten als auch bezüglich der realisationsphonetischen Plausibilität der Rekonstruktionen: Für die tschechischen Plosive wurde eine laminare Verschlusslösung mit einer relativ langen Konstriktion gefunden. Einschränkend ist dem allerdings hinzuzufügen, daß uns die recht tiefe Position des Kiefers fragwürdig erschien.

## Diskussion und Ausblick

Mit Story & Titzes Ansatz lässt sich ein Smoothing der an und für sich ‚untopologischen‘ Repräsentation von Areafunktionen durch Hauptkomponenten erzwingen. Das geht allerdings zu Lasten der realisationsphonetischen Validität der artikulatorischen Rekonstruktionen, da das Modell nicht über die Benutzung der ersten beiden Hauptkomponenten hinaus erweiterbar ist. Abhilfe würde die mit dem Konzept der artikulatorischen Faser einhergehende ‚lokale Regionslinearisierung‘ (Atal. et. al., 1978) schaffen, diese wurde jedoch bisher nicht unternommen. Eine andere Möglichkeit bestünde in der Verwendung einer regulierten Hauptkomponentenanalyse<sup>[11]</sup>.

Die angesprochenen Probleme bestanden bei der Verwendung des Optimierungsansatzes auf das artikulatorische Modell von Maeda nicht. Die Ursachen dafür dürften darin liegen, dass die Verwendung von Kontrollparametern für jeweils nur einen Teil der zu generierenden Areafunktion das oben als ‚Topologizität‘ bezeichnete Problem vermeidet. Andererseits erschien uns aber die recht tiefe Position des Kiefers fragwürdig. Eine Erklärung hierfür könnte in ‚Überoptimierung‘ bestehen: Auch in der Arbeit von Badin et. al.<sup>[12]</sup> wurden beträchtlich höhere Fehler in der Rekonstruktion des ersten Formanten gefunden. Wir waren zwar in der Lage, auch im ersten Formanten eine beinahe perfekte

Passung zu erzielen, dies ging möglicherweise auf Kosten der erwähnten unrealistisch tiefen Kieferposition. Dem wäre wahrscheinlich durch Verwendung einer modifizierten Kostenfunktion, welche Fehler im ersten Formanten geringer als Fehler in den höheren Formanten gewichtet, beizukommen.

Im Moment haben wir uns auf die Rekonstruktion statischer Konfigurationen beschränkt. In Zukunft wäre die Rekonstruktion von Trajektorien im Rahmen des gewählten Optimierungsansatzes interessant. Hierfür würde die jeweils vorherige statische Rekonstruktion als Startbedingung für den nächsten Zeitpunkt benutzt werden. Um glatte Trajektorien zu erhalten, wären die Intervalle, in denen man den Modellparametern Variation erlaubt, im Vergleich zu den hier für die statischen Rekonstruktionen (also ab dem zweiten zu optimierenden Sample) gewählten Parametern zu verringern.

## Danksagung

Gefördert von der Deutschen Forschungsgemeinschaft (DFG), Projekt GWZ 4/8-1.1.

<sup>1</sup> Keating, P. & Lahiri, A. Fronted velars, palatized velars, and palatals. *Phonetica*, 50, pp. 73-101, 1993.

<sup>2</sup> Wakita, H. Direct estimation of the vocal-tract hape by inverse filtering of acoustic speech wave-forms. *IEEE Trans. Audio Electroacoustics*, vol. 21, no. 5, pp. 417-27, 1973.

<sup>3</sup> Atal, B.S., Chang, J.J., Mathews, M.V. & Tukey, J.W. Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique. *JASA*, vol. 63, pp. 1535-1555, 1978.

<sup>4</sup> B. Story, B. & I. R. Titze, I. R. Parametrization of vocal tract area funtions by empirical orthogonal modes. *Journal of Phonetics*, pp. 223-260, 1998.

<sup>5</sup> Metropolis, N., Rosenbluth, A., Rosenbluth, M, Teller, A. & Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, pp. 1087-1092, 1953.

<sup>6</sup> Kirkpatrick, S., G., Gelatt, C. D. Jr. & Vecchi, M.P. Optimization by simulated annealing. *Science*, 220(4598), pp. 671-680, 1983.

<sup>7</sup> Ingber, L. Simulated annealing: Praticce versus theory. *Mathematical Computer Modeling*. 18(11), pp. 29-57,1993.

<sup>8</sup> Sondhi, M. M. & Schroeter, J. A hybrid time frequency domain articulatory speech synthesizer. *IEEE Transactions on Acoustics, Speech and Signal Processing*. 35, pp. 955-967,1987.

<sup>9</sup> Maeda, S. Compensatory articulation during speech: evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. in: Hardcastle, J. & Marchal, A.: *Speech Production and Speech Modeling*, Dordrecht: Kluwer Academic Publishers, 1990.

<sup>10</sup> *Handbook of the International Phonetic Association. A user's guide to the use of the International Phonetic Alphabet*. Cambridge,UK: Cambridge University Press, 1999.

<sup>11</sup> Ramsay, J. O. & Silverman, B. W. *Functional Data Analysis*. New York, Berlin: Springer, 1997.

<sup>12</sup> Badin, P., Beautemps, D., Laboissière, R. & Schwartz, J.L. Recovery of vocal tract geometry from formants for vowels and fricative consonants using a midsagittal-to-area function conversion model. *Journal of Phonetics*, 23, pp. 221-229, 1995.