

Titel: "VocalTractLab" als adäquates Entdeckungswerkzeug für Phonetik und Phonologie  
Name: Eva Lasarczyk, Universität des Saarlandes

Abstract:

Mit der Arbeit von Peter Birkholz (2005) wurde in neuerer Zeit ein (weiteres) System für artikulatorische Sprachsynthese verfügbar gemacht. Dieses "VocalTractLab" besteht aus einem dreidimensionalen Modell des Vokaltraktes, von der Lunge bis zu den Lippen. Mittels aerodynamisch-akustischer Simulation wird der Sprechschall durch die oralen und nasalen Höhlen und Nasennebenhöhlen geschickt und synthetische Sprache generiert.

Prinzipiell ist die Idee der artikulatorischen Sprachsynthese, dass sie, im Gegensatz zur z. B. Unit-Selection-Synthese, einen Einblick in die Sprachproduktion gewährleisten kann: Der geometrische Artikulationsprozess wird nicht nur als Black-Box betrachtet. Es werden auch nicht Sprachkorpusbausteine benutzt, deren Signaloberfläche geschickt manipuliert wird, um sprachliche Vielfalt zu erzeugen. Vielmehr kann direkt an artikulatorischen Stellschrauben gedreht werden. Die Frage ist nun, wie geeignet, angemessen und zuverlässig dieses Modell ist, wenn es um Grundlagenforschung geht. Kann man es als phonetisch-phonologisches "Entdeckungswerkzeug" einsetzen?

Auf dem Weg zu einer ersten Antwort auf diese Frage möchten wir eine Reihe von Experimenten präsentieren, die sich grob aufteilen lässt in supraglottale sowie glottale/subglottale Bereiche der Sprachproduktion. Desweiteren unterscheiden sie sich darin, in welchem Maße es sich um langfristige (relativ statische) Veränderungen handelt oder um kurzzeitige (eher segmentale) Manipulationen. Allen gemein ist, dass sie sich auf paralinguistische Gebiete konzentrieren und prinzipiell durch Perzeptionstests überprüft werden.

Bezüglich *langfristiger* Veränderungen bzw. Grundeinstellungen zeigt ein Experiment im glottalen Bereich zu Stimmqualität und Larynxhöhe, dass das Modell granular genug ist, um verschiedene Stimmqualitäten zu modellieren. Experimente im supraglottalen Bereich zeigen, dass auch langfristige Veränderungen (supraglottale "Settings") von z. B. den Lippen von Hörern erkannt werden. In dieser Richtung soll der Frage nachgegangen werden, ob sich verschiedene regionale Akzente auch durch unterschiedliche Langzeit-Settings erklären ließen.

In Bezug auf *kurzzeitige* Veränderungen zeigt ein Experiment im glottalen Bereich zu synthetischen Lachern, dass die Sprachsynthese vielversprechende Modellierungsmöglichkeiten bietet, deren Wirkungen auf den Hörer als positiv bewertet werden können. Ein Experiment im (supra)glottalen Bereich zu Sprechlachen (Sprache und Lachen gleichzeitig) hat aufgezeigt, dass auch hier eine perzeptive Wirkung erzielt werden kann. Ein weiteres Experiment in diesem Bereich - zu gelächelter Sprache - ergänzt die vorherigen Erkenntnisse.

Insgesamt haben die bisherigen Ergebnisse gezeigt, dass das Modell aufgrund seiner feinen Steuerungsmöglichkeiten und der guten Schallsimulation prinzipiell für die Grundlagenforschung geeignet zu sein scheint. Es wurde aber auch deutlich, an welchen Stellen noch Optimierungsbedarf besteht und dass ein solches System mit Bedacht eingesetzt werden sollte, um sich nicht von der natürlichen Sprachproduktion zu entfernen und unnatürliche Artikulationsbewegungen zu generieren. Insgesamt sind wir noch ein gutes Stück davon entfernt, dass artikulatorische Sprachsynthese ihre ganzen theoretischen Vorteile ausspielen kann (siehe Shadle & Damper, 2001).

Birkholz, P. 2005. 3D-Artikulatorische Sprachsynthese. University of Rostock (PhD thesis)

Shadle, C. H. and Damper, R. I. 2001. Prospects for articulatory synthesis: A position paper. In Proceedings of 4th ISCA Workshop on Speech Synthesis, August/September 2001, pp. 121-126, Pitlochry, Scotland.