

# An Articulatory and Phonatory Synthesis Model for Production of High Quality Speech and Singing

Peter Birkholz<sup>1</sup> and Bernd J. Kröger<sup>2</sup> (<http://www.speechtrainer.eu>)

<sup>1</sup>Institute for Computer Science, University of Rostock

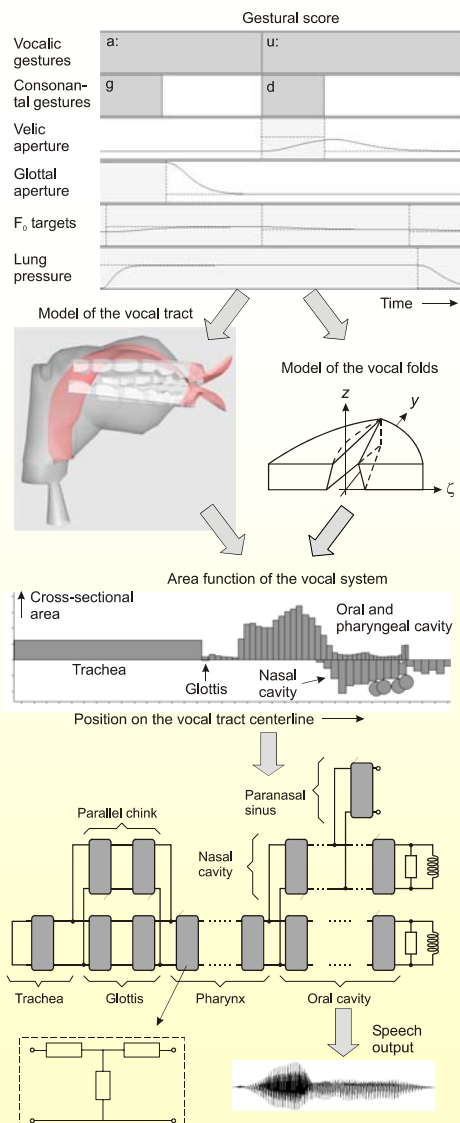
<sup>2</sup>Dept. of Phoniatics, Pedaudiology, and Communication Disorders, University Hospital Aachen and Aachen University

## Introduction

- Software implementation of a full parametric production model (synthesizer) including control modules for speech and singing
- Control units: articulatory and phonatory gestures (Birkholz et al. 2006)
- Articulatory-acoustic system: lumped element transmission line circuit including correction for acoustic and aerodynamic losses (Birkholz et al. 2007)
- Geometrical part of the glottis model after Titze (1984) and Cranen and Schroeter (1995)
- Direct control of glottal area over time (Birkholz 2005)
- Geometrical model of nasal tract including paranasal sinuses after Dang and Honda (1994 and 1996)

## The Synthesizer

Fig. 1. Overview of the synthesizer. Input: gestural score, calculation of articulatory movements and phonatory maneuvers; calculation of area function; calculation of air flow and pressure within each tube section; calculation of radiated sound wave.



## Rule-Based Generation of Gestural Scores

Fig. 2. Example of xml formatted gestural control sequences for "dona nobis pacem"

```
<song octaveOffset="0">
<note beatsPerMinute="110" pitch="rest" type="1/2" vibrato="0.5" lyrics="" loudness="1.0"
whisper="0" />
<note pitch="g3" type="1/8" lyrics="d o:" />
<note pitch="d3" type="1/8" lyrics="o:" />
<note pitch="h3" type="1/2" lyrics="n a:" />
<note pitch="a3" type="1/8" lyrics="n o:" />
<note pitch="d3" type="1/8" lyrics="o:" />
<note pitch="c4" type="1/2" lyrics="b i: s" />
....
</song>
```

## Pitch-Dependent Vocal Tract Target Shapes for Vowels

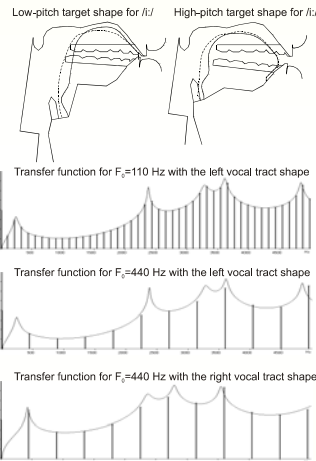
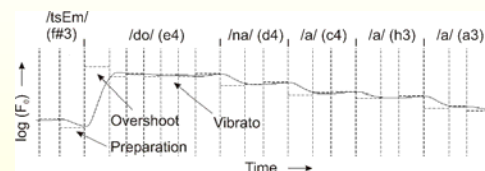


Fig. 3. Low pitch (left) and high pitch target shape (right) for the vowel /i:/ and the corresponding spectra for F<sub>0</sub>=110Hz and F<sub>0</sub>=440Hz. For the case F<sub>0</sub>=440Hz the spectrum is given for the low pitch target shape (false) and for the high pitch target shape (right).

Vowels at higher pitches are synthesized with a more "open" articulation. The low-pitch vocalic shapes are just adopted from speech articulation.

## Generation of F<sub>0</sub>-Contours

Fig. 4. F<sub>0</sub> contour and F<sub>0</sub> targets for "...dona...".



Each F<sub>0</sub>-gesture for realizing a note comprises a "preparation-phase", an "overshoot-phase" and the "target phase". During the target phase the F<sub>0</sub>-contour in addition exhibits vibrato.

## References

- Birkholz P (2005) 3D-Artikulatorische Sprachsynthese. Unpublished Ph.D.-thesis, University of Rostock, Germany
- Birkholz P, Jackel D, Kröger BJ (2006) Construction and control of a three-dimensional vocal tract model. *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2006)*, Toulouse, France, pp. 873-876
- Birkholz P, Jackel D, Kröger BJ (2007) Simulation of losses due to turbulence in the time-varying vocal system. *IEEE Transactions on Audio, Speech, and Language Processing* 15: 1218-1225
- Cranen B, Schroeter J (1995) Modeling a leaky glottis. *Journal of Phonetics* 23: 165-177
- Dang J, Honda K (1994) Morphological and acoustical analysis of the nasal and the paranasal cavities. *Journal of the Acoustical Society of America* 96: 2088-2100
- Dang J, Honda K (1996) Acoustical characteristics of the human paranasal sinuses derived from transmission characteristic measurement and morphological observation. *Journal of the Acoustical Society of America* 100: 3374-3383
- Titze IR (1984) Parameterization of the glottal area, glottal flow, and vocal fold contact area. *Journal of the Acoustical Society of America* 75: 570-580

More information concerning the software and for free download see: <http://www.vocaltractlab.de>