Articulation-to-Speech using Electro-Optical Stomatography and articulatory synthesis

TECHNISCHE UNIVERSITÄT DRESDEN

Simon Stone and Peter Birkholz

What is Articulation-to-Speech?

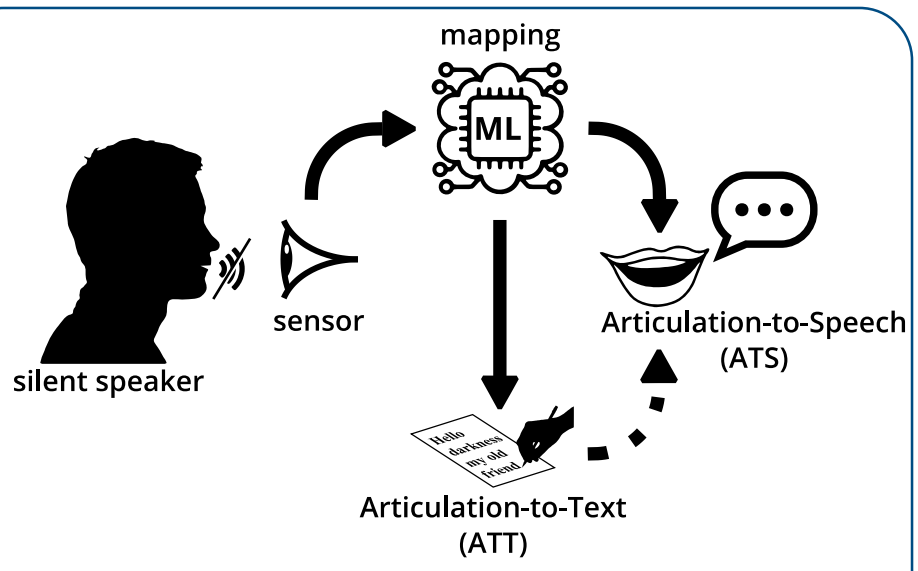

Figure 1: The concept of a Silent-Speech Interface

Why do we need Articulation-to-Speech?

Speaking aids:

- Increase volume or even correct articulation in the mapping step

Speech prostheses:

- Laryngectomees

Measurement technologies

- Electromyography (EMG) [1]
- Ultrasound [2]
- Permanent Magnetic Articulography (PMA) [3]

Electro-Optical Stomatography (EOS)

Novel technology developed by us

Intra-oral, pseudopalate-based

Combines and extends:

Electropalatography (EPG)

and

Optical Palatography (OPG)

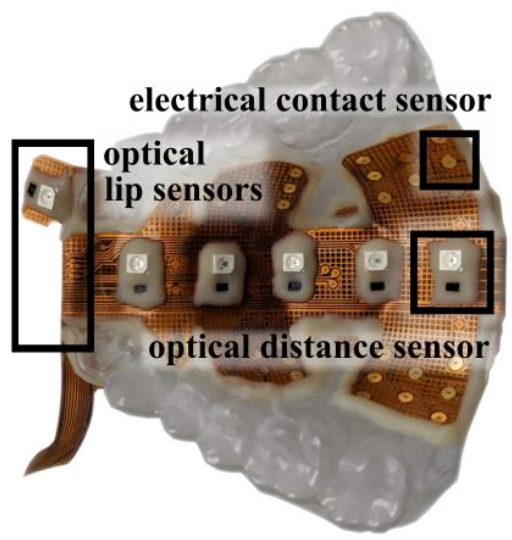

Figure 2: An EOS pseudopalate

Synthesis technologies for ATS

- Mostly mapping to MFCCs [1-3]
- Why not use articulatory data to control articulatory synthesizer?

Articulatory synthesis

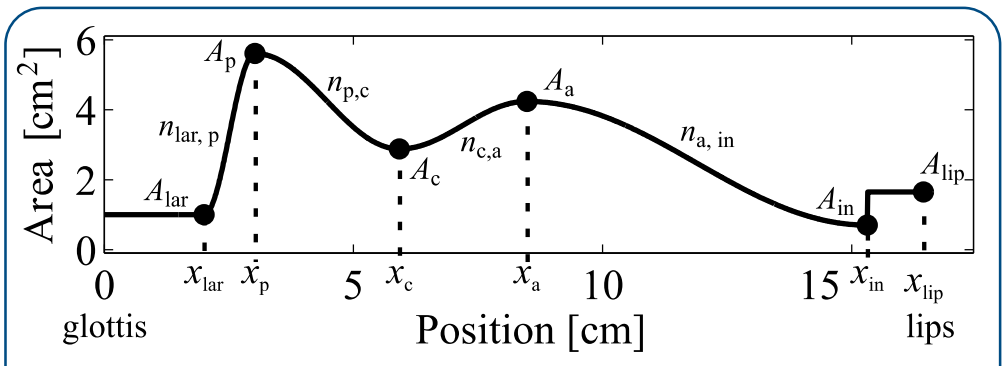

Figure 3: The six-point vocal tract model [4]

VocalTractLab acoustic synthesis backend to generate sound from vocal tract area function

Pilot study

- Four subjects recorded logatomes carrying all voiced German speech sounds in various contexts
- Trained mapping of median frames from each sound-of-interest to corresponding vocal tract shape

Table 1: Averaged 5-fold-cross-validated loss

	Least-Squares	SVM	Ensemble	GP
Subject 01	0.9794	0.8396	0.8226	0.8077
Subject 02	0.9378	0.8624	0.9378	0.7785
Subject 03	0.9313	0.8047	0.7276	0.9313
Subject 04	0.9078	0.6876	0.6783	0.6602

- Synthesis examples
- Vowels fairly recognizable
- Consonants unintelligible, likely due to
  - coarticulation,
  - ill-defined transitions

Outlook

- Generate coarticulated target vocal tract shapes
- Acoustic-to-articulatory inversion to obtain full trajectory in area function space
- Improve EOS measurement hardware

References

[1] M. Janke and L. Diener, "EMG-to-speech: Direct generation of speech from facial electromyographic signals," IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP), vol. 25, no. 12, pp. 2375–2385, 2017.

[2] T. G. Csapó, M. S. Al-Radhi, G. Németh, G. Gosztolya, T. Grósz, L. Tóth, and A. Markó, "Ultrasound-Based Silent Speech Interface Built on a Continuous Vocoder," in Proc. Interspeech 2019, 2019, pp. 894–898.

[3] J. A. Gonzalez, L. A. Cheah, A. M. Gomez, P. D. Green, J. M. Gilbert, S. R. Ell, R. K. Moore, and E. Holdsworth, "Direct speech reconstruction from articulatory sensor data by machine learning," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 25, no. 12, pp. 2362–2374.

[4] S. Stone, M. Marxen, and P. Birkholz. "Construction and evaluation of a parametric one-dimensional vocal tract model." IEEE/ACM Transactions on Audio, Speech, and Language Processing 26, no. 8 (2018): 1381-1392.