Evoc-Learn — High quality simulation of early vocal learning

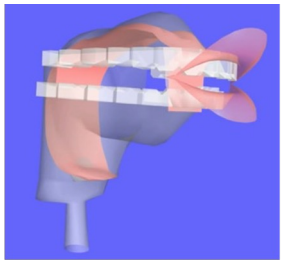
Yi Xu, Anqi Xu, Daniel R. van Niekerk, Branislav Gerazov, Peter Birkholz, Paul Konstantin Krug, Santitham Prom-on, Lorna F. Halliday

University College London, UK, UCMS, Skopje, RN Macedonia, Technische Universität, Dresden, Germany, King Mongkut's University of Technology Thonburi, Thailand, University of Cambridge, UK

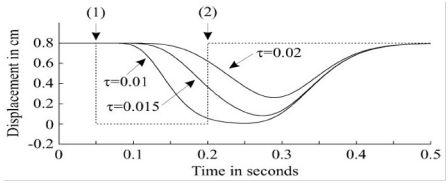
Outline

- Evoc-learn consists of tools for simulating vocal learning of highly intelligible speech
- Evoc-learn aims to resolves major bottlenecks in vocal learning:
  1. Lack of invariance
  2. Speaker normalization
- Components of Evoc-learn:
  - VocalTractLab — An articulatory synthesizer
  - A syllable-based coarticulation model
  - A sensory feedback system for guiding the learning process

VocalTractLab

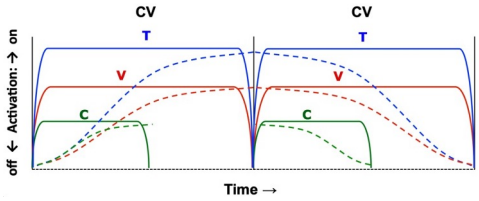


- 3-dimensional vocal tract
- Built-in aeroacoustic transformation



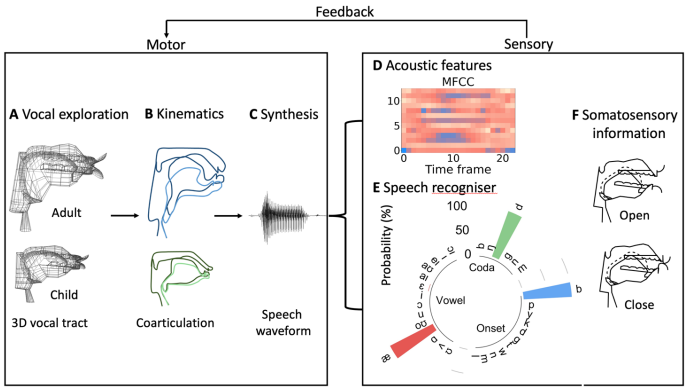
Built-in articulatory dynamics: Target Approximation

Syllable-based coarticulation



- Syllable is a mechanism for reducing temporal degrees of freedom by synchronizing consonantal (C), vocalic (V), and laryngeal (T) gestures at syllable onset.
- Target approximation is strictly sequential at the level of articulator dimensions, so that each dimensions can be only controlled by either consonant or vowel at any particular moment in time.
- This model helps to resolve much of the variability problem.

Sensory-guided vocal learning



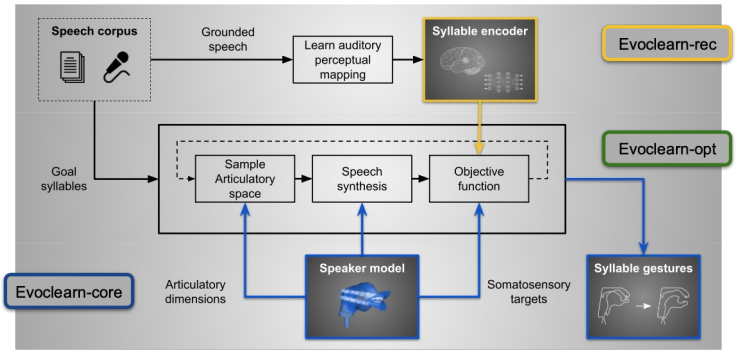
Four types of feedback mechanisms:

- Auditory matching — Simulated by acoustic fitting
- Perceptual recognition — Simulated by ASR
- Somatosensory constraints
- Visual feedback

For more information:

https://evoc-learn.gitlab.io

Software Components:



- A modular system implemented in Python as a number of standalone packages under the GNU General Public Licence
- Designed as sets of composable functional components which can be used to construct flexible processing pipelines for experiments either in Python directly or on the UNIX command line.
- **EVL-rec:** Construct and package standardised syllable encoders based on automatic speech recognition (ASR)
- **EVL-opt:** Goal-directed babbling as an optimisation process, including critical components for evaluation and feedback
- **EVL-core:** A thin layer that manages Pandas/Numpy, TA process, VocalTractLab

Funded by LEVERHULME TRUST